86

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
25 April 2002 (25.04.2002)

PCT

(10) International Publication Number
WO 02/33046 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number: PCT/US01/49451

(22) International Filing Date: 22 October 2001 (22.10.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/242,379 20 October 2000 (20.10.2000) US

(71) Applicant *(for all designated States except US)*: **UNIVERSITY OF MASSACHUSETTS** [US/US]; 1 Beacon Street, 26th Floor, Boston, MA 02108 (US).

(72) **Inventors; and**
(75) **Inventors/Applicants** *(for US only)*: **CZECH, Michael, P.** [US/US]; 75 Ruggles Street, Westborough, MA 01581 (US). **CHERNIACK, Andrew, D.** [US/US]; 22 Wigwam Hill Drive, Worcester, MA 01605 (US). **GUILHERME, Adilson, L.** [BR/US]; 46 Shrewsbury Green Drive, Apartment B, Shrewsbury, MA 01545 (US).

(74) **Agent: FASSE, Peter, J.**; Fish & Richardson P.C., 225 Franklin Street, Boston, MA 02110-2804 (US).

(81) **Designated States** *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) **Designated States** *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian

[Continued on next page]

(54) Title: GLUCOSE TRANSPORT-RELATED GENES AND USES THEREOF

(57) Abstract: Nucleotide sequences and amino acid sequences from nucleic acids and proteins involved in glucose transport are disclosed. The sequences are useful for producing DNA arrays that can be used for the diagnosis of, predictive testing for, and development of treatments for disorders involving glucose transport such as type II diabetes.

Novel Sequences from Clones in the Muscle-Adipocyte Union Library

Line number 259

>c0148

CCCCAACCTGCTCCATTGCTTGGGGGAGCGGTCCATGAGCGCTTGTCTCATCCCT
GGCCTCCCGGGAAAGTCTATGCAAAAGCTAAGGTTAACA (SEQ ID NO:1)

Line number 258

>c0827

TCACAGAGGCTCTGAGGCTACCACGAAGATGAACTCTCAGAAATGGGATTGTCA
CCCTCGATGAGTTTCCAGTTCCCTCTCTGTTGTATGATGACACAAGAAGGTGAAG
TGTTGCCTCTCTACAACTGGAAGAGGGAGA (SEQ ID NO:2)

Line number 260

>c1083

GTACTAGCGCTTACAGGTCTGTGTGCAGCCATGCCCAGCTTTCTAAGTGGGTGCT
GGAATCTGACTTCAGGTTTCATGCTTGAGCAGCAAAGCCCTCTTACACAGAGCCA
TTTCGACAGTTCTGTGACTTAGGTAGACTCACATCTGTCAGGCTAGAATTTCCAA
AATTGAAAATGAATTCAAAGTGAAATGCTTGGGAAGTAAGTTAAAGATAGGCTA
AATGGTTAACCCAGCAGTTAGGGTTGCTTTCTGCTCTTCCAGGGGACCTAAGTTT
GGTATTTTCTAGCACCCACACTTGGTGGCTCAAGCCCTCTAACTCCAGCTGCAGG
GGATAGGATGCCCCTCTTCTAACCTCCACTGGTGAGAATATCTCACACACACACAC
ACACACACACATGCTCACATACACAACCT (SEQ ID NO:3)

WO 02/33046 A2

patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

GLUCOSE TRANSPORT-RELATED GENES AND USES THEREOF

TECHNICAL FIELD

This invention relates to molecular biology, cell biology, glucose transport, medicine, and type II diabetes.

STATEMENT AS TO FEDERALLY SPONSORED RESEARCH

Work on this invention was supported in part with funds from the Federal government. The government therefore has certain rights in the invention.

BACKGROUND

Insulin stimulates glucose transport in muscle and fat. One of the most critical pathways that insulin activates is the rapid uptake of glucose from the circulation in both muscle and adipose tissue. Most of insulin's effect on glucose uptake in these tissues is dependent on the insulin-sensitive glucose transporter, GLUT4 (reviewed in Czech and Corvera, 1999, J. Biol. Chem. 274:1865-1868, Martin et al., 1999, Cell Biochem. Biophys. 30:89-113, Elmendorf et al., 1999 Exp. Cell Res. 253:55-62). The mechanism of insulin action is impaired in diabetes, leading to less glucose transport into muscle and fat. This is thought to be a primary defect in type II diabetes. Potentiating insulin action has a beneficial effect on type II diabetes. This is believed to be the mechanism of action of the drug Rezulin (troglitazone).

Type II diabetes mellitus (non-insulin-dependent diabetes) is a group of disorders, characterized by hyperglycemia that can involve an impaired insulin secretory response to glucose and insulin resistance. One effect observed in type II diabetes is a decreased effectiveness of insulin in stimulating glucose uptake by skeletal muscle. Type II diabetes

accounts for about 85-90% of all diabetes cases. In some cases of type II diabetes the underlying physiological defect appears to be multifactoral.

SUMMARY

The invention is based on the discovery of hundreds of genes that are preferentially expressed in cell types in which glucose transport is affected in type II diabetes, i.e., skeletal muscle and adipose tissue, as well as certain proteins expressed in glucose-transporting vesicles. Accordingly, the invention features methods of identifying a gene whose expression is altered in a glucose transport-related disease or disorder such as type II diabetes.

The invention includes a method of identifying a gene whose expression is altered in a glucose transport-related disorder. The method includes the steps of providing a nucleic acid array having 4 or more nucleic acids immobilized on a solid support, each nucleic acid having a sequence of 10 or more consecutive nucleotides within any one of the sequences listed in Figs. 1, 2A-2R, 3A-3E, 6A-6E, 7A-7U, 8A-8I, 9, 13A-13C, and 14A-14G or a complement thereof; providing a reference nucleic acid sample prepared from a tissue of a normal, control mammal; contacting the array with the reference sample; detecting hybridization of the reference sample with nucleic acids in the array, to obtain a reference pattern of glucose transport-related gene expression; providing a test nucleic acid prepared from a tissue of a mammal having a glucose transport-related disorder; contacting the array with the test sample; detecting hybridization of the test nucleic acid with nucleic acids in the array, to obtain a test pattern of glucose transport-related gene expression; and comparing the reference pattern with the test pattern to detect a gene whose expression is altered in the test pattern relative to its expression in the reference pattern. Figs. 6A-6E, 7A-7U, 8A-8I, 9, 13A-13C, and 14A-14G provide GenBank accession numbers. By accessing the sites indicated by the accession numbers, one in the art can obtain the nucleotide sequence and polypeptide sequence for the listed gene. In some embodiments, the array has 10 or more nucleic acids. In other embodiments, the array has 100 or more nucleic acids. In yet other embodiments, the array has not more than 100 nucleic acids, or not more than 300 nucleic acids. In certain embodiments of the invention, the sequence is 30 or more nucleotides in

length. The reference nucleic acid and the test nucleic acid can be cDNAs, that are, in some embodiments, fluorescently labeled.

The invention includes a nucleic acid array having 4 or more nucleic acids immobilized on a solid support, each nucleic acid having a sequence of 10 or more consecutive nucleotides within any one of sequences listed in Figs. 1, 2A-2R, 3A-3E, 6A-6E, 7A-7U, 8A-8I, 9, 13A-13C, and 14A-14G. In some embodiments, the array has 100 or more nucleic acids. In other embodiments, the array has not more than 100 nucleic acids, not more than 200 nucleic acids, or not more than 300 nucleic acids.

One aspect of the invention is an isolated nucleic acid molecule having a nucleotide sequence from any one of SEQ ID NOS:1-3, or a complement thereof. In some embodiments of the invention, the isolated nucleic acid sequence has a non-nucleic acid modifying group bound to either a 3' or 5' end of the nucleotide sequence or both; or a synthetic nucleic acid sequence bound to a 3' or 5' end of the nucleic acid sequence or both.

The invention also includes an isolated polypeptide having an amino acid sequence encoded by a nucleic acid sequence from any one of SEQ ID NOS:1-3.

Another embodiment of the invention is an isolated nucleic acid molecule having a nucleic acid sequence from any one of SEQ ID NOS:4-93, or a complement thereof. In certain embodiments, the nucleotide sequence has a non-nucleic acid modifying group bound to either a 3' or 5' end of the nucleotide sequence or both; or a synthetic nucleic acid sequence bound to a 3' or 5' end of the nucleic acid sequence or both. The invention includes an isolated nucleic acid molecule having a nucleic acid sequence selected from SEQ ID NOS:4-93, or a complement thereof. The invention also includes an isolated polypeptide having an amino acid sequence encoded by a nucleic acid sequence selected from any one of SEQ ID NOS:4-93.

In one aspect, the invention is method for identifying a candidate agent, that modulates the expression or activity of a glucose transport-related polypeptide. The method includes the steps of providing a sample containing a glucose transport-related polypeptide; adding a test agent to the sample; assaying the sample for expression or activity of the glucose transport-related polypeptide; and comparing the effect of the test agent on expression or activity of the glucose transport-related polypeptide relative to a control. A change in glucose transport-related polypeptide expression or activity indicates that the test

agent is a candidate agent that can modulate expression or activity of the glucose transport-related polypeptide. In some aspects of the method the test agent is a polynucleotide, a polypeptide, a small non-nucleic acid organic molecule, a small inorganic molecule, an antibody, an antisense oligonucleotide, or a ribozyme. In yet another embodiment, the glucose transport-related polypeptide is assayed using an antibody. In some embodiments of the invention, the glucose transport-related polypeptide is a human glucose transport-related polypeptide. The method can include the additional step of determining whether glucose transport is modulated in the presence of the test agent. The test agent can decrease or increase glucose transport. The assay can be a cell based assay or a cell-free assay. In certain embodiments of the invention, the glucose transport-related polypeptide is selected from the group of polypeptides encoded by sequences having the nucleic acid sequences listed in Figs. 1, 2A-2R, and 3A-3E, and the polypeptides listed in Figs. 6A-6E, 7A-7U, 8A-8I, 9, 13A-13C, and 14A-14G 6-9.

Modulation of expression (nucleic acid or polypeptide) or activity can be an increase or a decrease in expression or activity compared to a reference. The amount of modulation is generally at least two fold (i.e., a two fold increase or decrease in expression or activity) compared to a reference or a control sample. For example, the amount of modulation can be five fold, ten fold, fifty fold, 100 fold, or more.

The invention includes a method for identifying a candidate agent that modulates expression of a glucose transport-related polynucleotide. The method includes the steps of providing a sample in which a glucose transport-related polynucleotide is expressed; adding a test agent to the sample; detecting expression of the glucose transport-related polynucleotide; determining the amount of expression of the glucose transport-related polynucleotide; and comparing the effect of the test agent on the amount of expression of the glucose transport-related polynucleotide in the sample relative to a control, such that a change in the amount of expression from the glucose transport-related polynucleotide indicates the test agent is a candidate agent that can modulate expression of the glucose transport-related polynucleotide. The test agent can be a polynucleotide, a polypeptide, a small non-nucleic acid organic molecule, a small inorganic molecule, an antibody, an antisense oligonucleotide or a ribozyme. In some embodiments, the glucose transport-related polynucleotide is a human glucose transport-related polynucleotide. In another aspect of the invention, the method

includes the step of determining whether glucose transport is modulated (e.g., increased or decreased) in the presence of the test agent. In some embodiments, the glucose transport-related polynucleotide is selected from the group of sequences listed in Figs. 1, 2A-2R, and 3A-3E-3 or a complement thereof, and listed in Figs. 6A-6E, 7A-7U, 8A-8I, 9, 13A-13C, and 14A-14G, or a complement thereof. The assay used in the method can be cell-based assay or a cell-free assay.

The invention includes a method of diagnosing an individual having or at risk for a glucose transport-related disorder. The method includes the steps of providing a nucleic acid array having 4 or more nucleic acids immobilized on a solid support, each nucleic acid having a sequence of 10 or more nucleotides, the sequence having or containing a sequence selected from the group of the sequences listed in Figs. 1, 2A-2R, and 3A-3E, or a complement thereof, and the sequences of the genes listed in Figs. Figs. 6A-6E, 7A-7U, 8A-8I, 9, 13A-13C, and 14A-14G, or a complement thereof; providing a nucleic acid sample from the individual; contacting the array with the sample from the individual; detecting hybridization of nucleic acid in the sample from the individual with each nucleic acid in the array, to obtain a pattern of glucose transport-related gene expression; comparing the pattern of glucose transport-related gene expression in sample from the individual with a reference pattern, such that a comparison of the pattern of expression in the individual compared to the reference pattern indicates whether the individual has or is at risk for a glucose transport-related disorder. In some aspects of the invention, the array has 10 or more nucleic acids; or 100 or more nucleic acids. In other aspects of the invention, the array has not more than 100 nucleic acids; not more than 200 nucleic acids, or not more than 300 nucleic acids. In some embodiments, the sequence has 30 or more nucleotides. The sample from the individual can be a cDNA sample, and the cDNA sample can be fluorescently labeled. In some embodiments, the disorder is type II diabetes.

The invention also includes a nucleic acid array having 4 or more nucleic acids immobilized on a solid support, each nucleic acid comprising a sequence of 10 or more nucleotides, the sequence consisting of at least a portion of a sequence selected from the sequences listed in Figs. 1, 2A-2R, and 3A-3E, or a complement thereof, Figs. 6A-6E, 7A-7U, 8A-8I, 9, 13A-13C, and 14A-14G, or a complement thereof.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

Other features and advantages of the invention will be apparent from the detailed description, and from the claims.

DESCRIPTION OF DRAWINGS

Fig. 1 is a depiction of nucleic acid sequences identified in the Muscle Adipocyte Union library; c0148 (SEQ ID NO:1), c0827 (SEQ ID NO:2), and c1083 (SEQ ID NO:3).

Figs. 2A-2R are a series of sequences identified in the Muscle-Adipocyte Union Library (MAU library) that contain previously unidentified sequences and ESTs.

Figs. 3A-3E are series of sequences identified in the Adipocyte Subtractive (subtractive) library that contain previously unidentified sequences and ESTs.

Fig. 4 is a diagram showing a suppression subtractive hybridization protocol.

Fig. 5 is a diagram showing a protocol for constructing the Muscle-Adipocyte Union library.

Figs. 6A-6E are a table showing genes expressed in the Adipocyte Subtractive Library.

Figs. 7A-7U are a table showing genes expressed in the Muscle-Adipocyte Union Library.

Figs. 8A-8I are a table showing the proteins identified in peaks 1 and 2 of GLUT4-associated vesicles.

Fig. 9 is a table listing those proteins/genes that are present in one or both of the subtractive and Muscle-Adipocyte-Union libraries and were also identified as proteins purified from Glut4 vesicles. "Yes" indicates that a peptide(s) corresponding to the protein was present in a preparation. "?" indicates that the protein has not yet been identified in this preparation but its presence has not been excluded.

Figs. 10A-10D are a series of hydrophobicity plots of the c0582 sequence.

Figs. 11A-11D are a series of hydrophobicity plots of the c0139 sequence.

Figs. 12A-12D are a series of hydrophobicity plots of the b0175 sequence.

Figs. 13A-13C are a table listing genes whose expression was not detected in fibroblasts, and was detected in adipocyte or muscle using GeneChips. Columns marked f1 and f2 are data from the fibroblast replicate chips, columns marked a1 and a2 are data from the adipocyte replicate chips, and the columns marked m1 and m2 are data from the muscle replicate chips. A indicates that the gene is absent in a tissue. P indicates that the gene is present in a tissue. An M indicates marginal signal and the software cannot determine if the gene is absent or present.

Figs. 14A-14G are tables listing genes whose expression was determined to be the same on all fibroblast chips, and increased on both adipocyte or muscle GeneChips compared to a fibroblast chip. The columns marked f1, f2, and f3 are fibroblast replicate chips. The columns marked a1, a2, and a3 are adipocyte replicate chips, and the columns marked m1, m2, and m3 are the muscle replicate chips. NC indicates no change of expression. MI indicates that there was a moderate increase in expression. An I indicates an increase in expression. The function classes of the genes listed in the last column are as follows: Class 1 genes encode metabolic proteins; Class 2 genes encode signaling proteins.

Figs. 15A-15B are a table listing highly expressed genes common between the Muscle-Adipocyte Union library and the Mu-74 GeneChips Arrays.

DETAILED DESCRIPTION

Library of Glucose Transport-Related Sequences

Suppressive subtraction hybridization has been applied to create libraries (databases) of glucose transport-related nucleotide sequences. The Muscle-Adipocyte Union library contains about 230 glucose transport-related nucleotide sequences and was made by identifying nucleotide sequences selectively expressed in fat and muscle tissue, but not in fibroblasts. Sequences from the subtractive library or the MAU library can be used in the invention. Generally, the sequences are from the MAU library. Unless indicated otherwise below, the library referred to is the MAU library. The sequences in the library represent glucose transport-related genes that are candidates for involvement in insulin-related action.

and thus potential drug targets for glucose transport-related disorders. Glucose transport-related disorders include diseases such as type II diabetes, obesity, certain types of cardiovascular disease, and Syndrome X.

The library can be used to construct DNA arrays for identifying glucose transport-related genes whose expression is altered (increased or decreased) in diseases or disorders characterized by insulin resistance, e.g., type II diabetes, or defects in glucose transport. The library advantageously enables gene expression pattern comparisons that involve tens or hundreds of genes most likely to be involved in insulin resistance and type II diabetes, instead of comparisons that involve tens of thousands or hundreds of thousands of genes. This focus on a relatively small library advantageously simplifies data analysis and improves the signal-to-noise ratio. In addition to being useful for identifying individual glucose transport-related genes, DNA arrays of the invention can be used to identify gene expression patterns indicative of particular forms of type II diabetes or a predisposition (i.e., at risk for) for development of type II diabetes. The predisposition can be a genetic predisposition.

Once specific glucose transport-related genes are identified using the library, assays for expression of individual genes can be employed. Specific assays can be employed, for example, in diagnostic methods to diagnose type II diabetes, methods for diagnosing particular forms of type II diabetes, and methods for identifying individuals who have pre-symptomatic forms of type II diabetes or a genetic predisposition for development of type II diabetes. Such diagnostic assays may provide useful information for devising therapeutic strategies tailored to individual patients.

The library can also be used to assay expression of individual genes in animal (e.g., mouse) models of a disease in which glucose transport is affected. For example, cDNA can be prepared from RNA isolated from a mouse having a glucose transport-related disorder such as diabetes. The RNA can be isolated from a tissue that normally carries out glucose transport (e.g., muscle or adipose tissue). The cDNA is hybridized to sequences from the MAU library. Expression of the MAU library sequences is then compared to expression of the sequences in a mouse that does not have the disorder. A relative increase or decrease in the expression of a sequence in the mouse having a glucose transport disorder compared to an unaffected mouse indicates that the sequence is involved in the disorder. Such sequences are useful, e.g., for indicating genes or gene products as drug targets for treating the disorder.

Sequences in the MAU library fall into three categories: (1) novel sequences (Fig. 1); (2) sequences from genes for which at least partial sequences were known, but for which no function was known or predicted (Figs. 2A-2R and 3A-3E); and (3) sequences of genes with a known or predicted function (included in Figs. 6A-6E and 7A-7U). The novel sequences are designated c0148 (SEQ ID NO:1), c0827 (SEQ ID NO:2), and c1083 (SEQ ID NO:3), and they are set forth in Fig. 1.

Some of the library sequences are a novel combination of sequences based on partial sequencing of genes that were identified in the Adipocyte Subtractive library as differentially expressed in adipocyte and fibroblast cells combined with overlapping sequences that were obtained from databanks (GenBank and TIGR (The Institute for Genomic Research)). Additional library sequences are novel combinations of sequences based on partial sequencing of genes that are identified in the Muscle Adipocyte Union Library as differentially expressed in both adipocyte and muscle cells combined with overlapping sequences that were obtained from the databanks. Genes in these categories include b0117 (AAPT-like protein with CDP-alcohol phosphatidyltransferases signature sequence; SEQ ID NO:81), b0175 (GS2 protein; SEQ ID NO:87), c0139 (endophilin-like protein coil-coil plus SH3 domain; SEQ ID NO:12), c0250 (SEQ ID NO:17), c0352 (SEQ ID NO:18), c0582 (Rab GTPase domain; SEQ ID NO:33), c0591 (isoform of TIG2 protein; SEQ ID NO:34), and c0840 (Clu-like protein; SEQ ID NO:53). These sequences are depicted in Figs. 2A-2R and 3A-3E, and are particularly useful in the methods of the invention.

Sequences that are differentially expressed in adipocytes, muscle cells, or both (as compared to expression in, e.g., fibroblasts) are useful, e.g., as genes or providing gene products that are targets for treatments for disorders involving glucose transport and for diagnosis of disorders involving aberrant glucose transport such as type II diabetes.

DNA Arrays

DNAs containing complete or partial sequences from the library of glucose transport-related sequences can be used to construct conventional DNA arrays (sometimes called DNA chips or gene chips). A DNA array according to the invention can contain tens, hundreds, or thousands of individual sequences immobilized (tethered) at discrete, predetermined locations (addresses or "spots") on a solid, planar support, e.g., glass or nylon. Each spot

may contain more than one DNA molecule, but each DNA molecule at a given address has an identical nucleotide sequence. The DNA array can be a macroarray or microarray, the difference being in the size of the DNA spots. Macroarrays contain spots of about 300 microns in diameter or larger and can be imaged using gel or blot scanners. Microarrays contain spots less than 300 microns, typically less than 200 microns, in diameter.

For analysis and comparison of glucose transport-related gene expression patterns, an array is constructed using sequences from at least four, e.g., at least 10, 20, 40, 60, 80 or 100 genes in the above-described library. A population of labeled cDNA representing total mRNA from a sample of a tissue of interest, e.g., muscle or adipose tissue, is contacted with the DNA array under suitable hybridization conditions. Hybridization of cDNAs with sequences in the array is detected, e.g., by fluorescence at particular addresses on the solid support. Thus, a pattern of fluorescence representing a gene expression pattern in the tissue of a particular individual or group of individuals is obtained. These patterns of glucose transport-related gene expression can be digitized and stored electronically for computerized analysis and comparison. For example, an array according to the invention can be used to compare glucose transport-related gene expression of type II diabetic individuals with each other, and with non-diabetic individuals. Such comparisons will reveal specific genes whose expression is increased or decreased in a given tissue type in individuals with type II diabetes or other glucose transport-related diseases or disorders. Such arrays can also be used to diagnose individuals having or at risk for a glucose transport-related disorder such as type II diabetes. For example, a nucleic acid sample (e.g., cDNA) from an individual suspected of having a glucose transport-related disorder is prepared and hybridized to the array. The pattern (including the level) of expression of sequences in the sample is compared to a reference pattern (e.g., representing the pattern of expression in unaffected individuals, and/or representing the pattern of expression in individuals known to have a particular glucose transport-related disorder). A pattern of expression in the sample that varies from that of the unaffected reference, and/or corresponds with the pattern of expression in a glucose transport disorder indicates that the individual has a glucose transport disorder.

In some embodiments of the invention, cDNAs are used to form the array. Suitable cDNAs can be obtained by conventional polymerase chain reaction (PCR) techniques. The length of the cDNAs can be from 20 to 2,000 nucleotides, e.g., from 100 to 1,000

nucleotides. Other methods known in the art for producing cDNAs can be used. For example, reverse transcription of a cloned sequence can be used (for example, as described in Sambrook et al., eds., Molecular Cloning: A Laboratory Manual. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press. Cold Spring Harbor, NY, 1989)

The cDNAs are placed ("printed" or "spotted") onto a suitable solid support (substrate), e.g., a coated glass microscope slide, at specific, predetermined locations (addresses) in a two-dimensional grid. A small volume, e.g., 5 nanoliters, of a concentrated DNA solution is used in each spot. Spotting can be carried out using a commercial microspotting device (sometimes called an arraying machine or gridding robot) according to the vendor's instructions. Commercial vendors of solid supports and equipment for producing DNA arrays include BioRobotics Ltd., Cambridge, UK: Corning Science Products Division, Acton, MA; GENPAK Inc., Stony Brook, NY; SciMatrix, Inc., Durham, NC; and TeleChem International, Sunnyvale, CA.

The cDNAs can be attached to the solid support by any suitable method. In general, the linkage is covalent. Suitable methods of covalently linking DNA molecules to the solid support include amino cross-linking and UV crosslinking. For guidance concerning construction of cDNA arrays according to the invention, see, e.g., DeRisi et al., 1996, *Nature Genetics* 14:457-460; Khan et al., 1999, *Electrophoresis* 20:223-229; Lockhart et al., 1996, *Nature Biotechnol.* 14:1675-1680.

In some embodiments of the invention, the immobilized DNAs in the array are synthetic oligonucleotides. Preformed oligonucleotides can be spotted to form a DNA array, using techniques described above with regard to cDNA. In general, however, the oligonucleotides are synthesized directly on the solid support. Methods for synthesizing oligonucleotide arrays are known in the art. See, e.g., Fodor et al., U.S. Patent No. 5,744,305. The sequences of the oligonucleotides represent portions of the sequences in the library described above. For example, the lengths of oligonucleotides are 10 to 50 nucleotides, e.g., 15, 20, 25, 30, 35, 40, or 45 nucleotides.

In some embodiments of the invention, the human homologs of the identified sequences are used in the detection method. Examples of such human homologs are listed with their GenBank accession numbers in Figs. 6A-6E, 7A-7U, and 8A-8I. In other embodiments, the sequence used for detection consists of highly conserved regions of the

sequence, e.g., sequence that is highly conserved between homologous mouse and human sequence.

Sample Preparation and Analysis

In methods of the invention, the transcription level of a glucose transport-related gene is assumed to be reflected in the amount of its corresponding mRNA present in cells of assayed tissue or cell lines derived from specific tissues. In general, mRNA from the cells or tissue is copied into cDNA under conditions such that the relative amounts of cDNA produced representing specific genes reflect the relative amounts of the mRNA in the sample. Comparative hybridization methods involve comparing the amounts of various, specific mRNAs in two tissue samples, as indicated by the amounts of corresponding cDNAs hybridized to sequences from the glucose transport-related gene library.

The mRNA used to produce cDNA is generally isolated from other cellular contents and components. One useful approach for mRNA isolation is a two-step approach. In the first step, total RNA is isolated. The second step is based on hybridization of the poly(A) tails of mRNAs to oligo(dT) molecules bound to a solid support, e.g., a chromatographic column or magnetic beads. Total RNA isolation and mRNA isolation are known in the art and can be accomplished, for example, using commercial kits according to the vendor's instructions. Similarly, synthesis of cDNA from isolated mRNA is known in the art and can be accomplished using commercial kits according to the vendor's instructions. Fluorescent labeling of cDNA can be achieved by including a fluorescently labeled deoxynucleotide, e.g., Cy5-dUTP or Cy3-dUTP, in the cDNA synthesis reaction. For guidance concerning isolation of mRNA and synthesis of fluorescently labeled cDNA for analysis on a DNA array, see, e.g., Ross et al., 2000, *Nature Genetics* 24:227-235.

In the invention, conventional techniques for hybridization and washing of DNA arrays, detection of hybridization, and data analysis can be employed routinely without undue experimentation. Commercial vendors of hardware and software for scanning DNA arrays and analyzing data include Cartesian Technologies, Inc. (Irvine, CA); GSI Lumonics (Watertown, MA); Genetic Microsystems Inc. (Woburn, MA); and Scanalytics, Inc. (Fairfax, VA).

Isolated Nucleic Acid Molecules

The invention provides certain novel, isolated nucleic acids that encode murine glucose transport-related polypeptides, or biologically active portions thereof (Fig. 1). In addition to forming part of the library, these nucleic acids can be used as hybridization probes to identify the full-length genes that they represent, and to isolate related nucleic acids, e.g., murine nucleic acids can be used to identify and clone human homologs. These nucleic acids also can be used to design PCR primers for PCR amplification of related nucleic acid molecules. The full-length genes identified and isolated using these novel sequences are predicted to function in insulin-responsive glucose transport systems in mammalian muscle cells and adipose cells.

As used herein, "isolated DNA" means DNA that has been separated from DNA that flanks the DNA in the genome of the organism in which the DNA naturally occurs. The term therefore includes recombinant DNA incorporated into a vector, e.g., a cloning vector or an expression vector. The term also includes a molecule such as a cDNA, a genomic fragment, a fragment produced by PCR, or a restriction fragment. The term also includes a recombinant nucleotide sequence that is part of a hybrid gene construct, i.e., a construct encoding a fusion protein. The term excludes an isolated chromosome. Isolated nucleic acids of the invention (e.g., SEQ ID NOS:1-93) can include modifications at the 3' and/or 5' end of the molecule including a metal, a modified nucleotide residue, or a nucleotide sequence that is not contiguous with the sequence of interest in nature. Such modifications can also be made to the sequences or fragments of sequences used in the invention (e.g., sequences derived from the genes listed in Figs. 6-9 and 13-15).

A full length coding sequence that contains a novel nucleotide sequence of the invention, e.g., a nucleic acid molecule containing a sequence set forth in Fig. 1, or a complement thereof, can be isolated using conventional molecular biology techniques and the sequence information provided herein. For example the isolation can be accomplished without undue experimentation by applying techniques described in numerous treatises and reference manuals. For general guidance and specific protocols, see, e.g., Sambrook et al., eds., *Molecular Cloning: A Laboratory Manual, 2nd ed.*, Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989; Ausubel et al. (eds.),

1994, *Current Protocols in Molecular Biology*, John Wiley & Sons, Inc.; Innes et al. (eds.), 1990, *PCR Protocols*, Academic Press.

A nucleic acid molecule of the invention can be amplified using cDNA, mRNA, or genomic DNA as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. Once isolated, the full-length nucleic acid can be cloned into an appropriate vector and characterized by conventional DNA sequence analysis, using standard techniques and equipment.

A nucleic acid fragment encoding a biologically active portion of a polypeptide encoded by a novel nucleic acid of the invention can be identified and prepared by isolating a portion of any of the sequences useful in the invention, expressing the encoded portion of the polypeptide protein (e.g., by recombinant expression *in vitro*) and assessing the activity of the encoded portion of the polypeptide.

The invention further encompasses nucleic acid molecules that differ from the nucleotide sequence set forth in Fig. 1, due to degeneracy of the genetic code and thus encode the same amino acid sequence as that encoded by the nucleotide sequence set forth in Fig. 1. The invention further encompasses isolated nucleic acid molecules that hybridize with the sequences set forth in Fig. 1 under high stringency conditions. As used herein, "high stringency" means the following: hybridization at 42° C in the presence of 50% formamide; a first wash at 65° C with 2 x SSC containing 1% SDS; followed by a second wash at 65° C with 0.1 x SSC.

In addition to the nucleotide sequences set forth in Fig. 1, it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequence may exist within a population (e.g., the human population). Such genetic polymorphisms may exist among individuals within a population due to natural allelic variation. An allele is one of a group of genes that occur alternatively at a given genetic locus. As used herein, "allelic variation" means variation in a nucleotide sequence that occurs at a given locus, or variation in an amino acid sequence of a polypeptide encoded by the nucleotide sequence at a given locus. Alternative alleles can be identified by sequencing the gene of interest in a number of different individuals. This can be accomplished by using hybridization probes to identify nucleic acids corresponding to the same genetic locus in a variety of individuals. The nucleic acid is then sequenced (e.g., amplified using PCR and the

PCR products are sequenced) to identify variations. Isolated nucleic acids containing the nucleotide sequences of Fig. 1 that display allelic variations while retaining functional activity are within the scope of the invention.

In some embodiments of the invention, changes are introduced into the sequences of Fig. 1 by mutation thereby leading to changes in the amino acid sequence of the encoded protein, without altering the biological activity of the protein. For example, one can make nucleotide substitutions leading to amino acid substitutions at non-essential amino acid residues. A non-essential amino acid residue is a residue that can be altered from the wild-type sequence without altering the biological activity of the gene product (e.g., a protein). For example, amino acid residues that are not conserved or only semi-conserved among homologs of various species may be non-essential for activity and thus would be likely targets for alteration. In contrast, amino acid residues that are conserved among the homologs of various species (e.g., murine and human) may be necessary for activity and thus would not be likely targets for alteration.

An isolated nucleic acid molecule encoding a variant protein can be created by introducing one or more nucleotide substitutions, additions, or deletions into the nucleotide sequence of c0148 (SEQ ID NO:1), c0827 (SEQ ID NO:2), and c1083 (SEQ ID NO:3) such that one or more amino acid substitutions, additions, or deletions are introduced into the encoded protein. Mutations can be introduced by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (e.g., lysine, arginine, histidine), acidic side chains (e.g., aspartic acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (e.g., threonine, valine, isoleucine) and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, histidine). Alternatively, mutations can be introduced randomly along all or part of the coding sequence, such as by saturation mutagenesis, and the resultant mutants can be

screened for biological activity to identify mutants that retain activity. Following mutagenesis, the encoded protein can be expressed recombinantly and the activity of the protein can be determined.

Isolating Homologous Sequences from Other Species

The human homologs of glucose-transport related genes and their products are useful for various embodiments of the present invention including diagnosis of glucose transport-related disorders such as type II diabetes. Homologs have already been identified for certain genes and GenBank Accession numbers are supplied for these. In those cases where a human homolog is not identified, several approaches can be used to identify such genes. These methods include low stringency hybridization screens of human libraries with a mouse glucose transport-related nucleic acid sequence, polymerase chain reactions (PCR) of human DNA sequence primed with degenerate oligonucleotides derived from a mouse glucose transport-related gene, two-hybrid screens, and database screens for homologous sequences.

Antisense Nucleic Acids

The invention includes antisense nucleic acid molecules, i.e., nucleic acid molecules whose nucleotide sequence is complementary to all or part of an mRNA based on the sequences c0148, c0827, and c1083 (Fig. 1). An antisense nucleic acid molecule can be antisense to all or part of a non-coding region of the coding strand of a nucleotide sequence encoding a polypeptide of the invention. The non-coding regions ("5' and 3' untranslated regions") are the 5' and 3' sequences that flank the coding region and are not translated into amino acids.

An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45, or 50 nucleotides or more in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (e.g., an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, e.g., phosphorothioate derivatives and acridine substituted nucleotides can be used. Examples of

modified nucleotides which can be used to generate the antisense nucleic acid include 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (i.e., RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention can be administered to a mammal, e.g., a human patient. Alternatively, they can be generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a selected polypeptide of the invention to thereby inhibit expression, e.g., by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarities to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule which binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention includes direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface, e.g., by linking the antisense nucleic acid molecules to peptides or antibodies that bind to cell surface receptors or antigens. The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. For example, to achieve sufficient intracellular concentrations of

the antisense molecules, vector constructs can be used in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter.

An antisense nucleic acid molecule of the invention can be an α-anomeric nucleic acid molecule. An α-anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual, β-units, the strands run parallel to each other (Gaultier et al., 1987, *Nucleic Acids Res.* 15:6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue et al., 1987, Nucleic Acids Res. 15:6131-6148) or a chimeric RNA-DNA analog (Inoue et al., 1987, FEBS Lett. 215:327-330).

Antisense molecules that are complementary to all or part of a glucose transport-related gene are also useful for assaying expression of such genes using hybridization methods known in the art. For example, the antisense molecule is labeled (e.g., with a radioactive molecule) and an excess amount of the labeled antisense molecule is hybridized to an RNA sample. Unhybridized labeled antisense molecule is removed (e.g., by washing) and the amount of hybridized antisense molecule measured. The amount of hybridized molecule is measured and used to calculate the amount of expression of the glucose transport-related gene. In general, antisense molecules used for this purpose can hybridize to a sequence from a glucose transport-related gene under high stringency conditions such as those described herein. When the RNA sample is first used to synthesize cDNA, a sense molecule can be used. It is also possible to use a double-stranded molecule in such assays as long as the double-stranded molecule is adequately denatured prior to hybridization.

Ribozymes

The invention also encompasses ribozymes that have specificity for the sequences c0148, c0827, and c1083. Ribozymes are catalytic RNA molecules with ribonuclease activity that are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (e.g., hammerhead ribozymes (described in Haselhoff and Gerlach, 1988, Nature 334:585-591)) can be used to catalytically cleave mRNA transcripts to thereby inhibit translation of the protein encoded by the mRNA. A ribozyme having specificity for a nucleic acid molecule of the invention can be designed based upon the nucleotide sequence of a cDNA disclosed herein. For example, a derivative

of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a glucose transport-related mRNA (Cech et al. U.S. Patent No. 4,987,071; and Cech et al., U.S. Patent No. 5,116,742). Alternatively, an mRNA encoding a polypeptide of the invention can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. *See, e.g.*, Bartel and Szostak, 1993, Science 261:1411-1418.

The invention also encompasses nucleic acid molecules that form triple helical structures. For example, expression of a polypeptide of the invention can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the gene encoding the polypeptide (e.g., the promoter and/or enhancer) to form triple helical structures that prevent transcription of the gene in target cells. *See generally* Helene, 1991, Anticancer Drug Des. 6(6):569-84; Helene, 1992, Ann. N.Y. Acad. Sci. 660:27-36; and Maher, 1992, Bioassays 14(12):807-15.

In various embodiments, the nucleic acid molecules of the invention can be modified at the base moiety, sugar moiety or phosphate backbone to improve, e.g., the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids (*see* Hyrup et al., 1996, Bioorganic & Medicinal Chemistry 4(1): 5-23). Peptide nucleic acids (PNAs) are nucleic acid mimics, e.g., DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs allows for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols, e.g., as described in Hyrup et al., 1996, *supra*; Perry-O'Keefe et al., 1996, *Proc. Natl. Acad. Sci. USA* 93: 14670-675.

PNAs can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, e.g., inducing transcription or translation arrest or inhibiting replication. PNAs can also be used, e.g., in the analysis of single base pair mutations in a gene by, e.g., PNA directed PCR clamping; as artificial restriction enzymes when used in combination with other enzymes, e.g., S1 nucleases (Hyrup, 1996, *supra*; or as probes or primers for DNA sequence

and hybridization (Hyrup, 1996, *supra*; Perry-O'Keefe et al., 1996, *Proc. Natl. Acad. Sci. USA* 93: 14670-675).

PNAs can be modified, e.g., to enhance their stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras can be generated which may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes, e.g., RNAse H and DNA polymerases, to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (Hyrup,1996, *supra*). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup,1996, *supra*, and Finn et al., 1996, *Nucleic Acids Res.* 24:3357-63. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry and modified nucleoside analogs. Compounds such as 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite can be used as a link between the PNA and the 5' end of DNA (Mag et al., 1989, *Nucleic Acids Res.* 17:5973-88). PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment (Finn et al., 1996, *Nucleic Acids Res.* 24:3357-63). Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment (Peterser et al., 1975, *Bioorganic Med. Chem. Lett.* 5:1119-11124).

In some embodiments, the oligonucleotide includes other appended groups such as peptides (e.g., for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (*see, e.g.*, Letsinger et al., 1989, *Proc. Natl. Acad. Sci. USA* 86:6553-6556; Lemaitre et al., 1987, *Proc. Natl. Acad. Sci. USA* 84:648-652; PCT Publication No. W0 88/09810) or the blood-brain barrier (*see, e.g.*, PCT Publication No. W0 89/10134). In addition, oligonucleotides can be modified with hybridization-triggered cleavage agents (*see, e.g.*, Krol et al., 1988, *BioTechniques* 6:958-976) or intercalating agents (*see, e.g.*, Zon, 1988, *Pharm. Res.* 5:539-549). To this end, the oligonucleotide may be conjugated to another molecule, e.g., a peptide, hybridization triggered cross-linking agent, transport agent, hybridization-triggered cleavage agent, etc.

Isolated Proteins

The invention provides isolated polypeptides encoded by glucose transport-related nucleic acids depicted in Figs. 1, 2A-2R, and 3A-3E. These polypeptides can be used, e.g., as immunogens to raise antibodies. Methods are well known in the art for predicting the translation products of the nucleic acids (i.e, using computer programs that provide the predicted polypeptide sequences and direction as to which of the three reading frames is the open reading frame of the sequence. These polypeptide sequences can then be produced either biologically (e.g., by positioning the nucleic acid sequence that encodes them in-frame in an expression vector transfected into a compatible expression system) or chemically using methods known in the art. The polypeptides encoded by the genes listed in Figs. 6-9 and 13-15 are also useful in the invention. For example, the entire polypeptide or a fragment thereof can be used to produce an antibody that is useful in a screening assay. Figs. 6-9 and 13-15, provide the GenBank accession numbers of the sequences, when available. These listings provide both nucleotide and polypeptide sequences that are useful in the invention.

An "isolated" or "purified" protein or biologically active portion thereof is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the protein is derived, or substantially free of chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of protein in which the protein is separated from cellular components of the cells from which it is isolated or recombinantly produced. Thus, protein that is substantially free of cellular material includes preparations of protein having less than about 30%, 20%, 10%, or 5% (by dry weight) of heterologous protein (also referred to herein as "contaminating protein"). In general, when the protein or biologically active portion thereof is recombinantly produced, it is also substantially free of culture medium, i.e., culture medium represents less than about 20%, 10%, or 5% of the volume of the protein preparation. In general, when the protein is produced by chemical synthesis, it is substantially free of chemical precursors or other chemicals, i.e., it is separated from chemical precursors or other chemicals that are involved in the synthesis of the protein. Accordingly such preparations of the protein have less than about 30%, 20%, 10%, 5% (by dry weight) of chemical precursors or compounds other than the polypeptide of interest.

Expression of proteins and polypeptides can be assayed to determine the amount of expression. Methods for assaying protein expression are known in the art and include Western blot, immunoprecipitation, and radioimmunoassay.

Biologically active portions of a polypeptide of the invention include polypeptides comprising amino acid sequences sufficiently identical to or derived from the amino acid sequence of the protein, which include fewer amino acids than the full length protein, and exhibit at least one activity of the corresponding full-length protein. Typically, biologically active portions comprise a domain or motif with at least one activity of the corresponding protein. A biologically active portion of a protein of the invention can be a polypeptide which is, for example, 10, 25, 50, 100, or more amino acids in length. Moreover, other biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of the native form of a polypeptide of the invention.

Polypeptides of the invention have the predicted amino acid sequence of an open reading frame of c0148 (SEQ ID NO:1), c0827 (SEQ ID NO:2), and c1083 (SEQ ID NO:3). In some embodiments, polypeptides of the invention have the predicted amino acid sequence selected from SEQ ID NOS:4-93. Other useful proteins are substantially identical (e.g., at least about 45%, preferably 55%, 65%, 75%, 85%, 95%, or 99%) to the predicted amino acid sequence of a polypeptide encoded by a polynucleotide comprising the polynucleotide sequence of c0148 (SEQ ID NO:1), c0827 (SEQ ID NO:2), and c1083 (SEQ ID NO:3) or substantially identical (e.g., at least about 93%, preferably 94%, 95%, 96%, or 99%) to the predicted amino acid sequence of a polypeptide encoded by a polynucleotide comprising the polynucleotide sequence of c0148 (SEQ ID NO:1), c0827 (SEQ ID NO:2), and c1083 (SEQ ID NO:3), and retain the functional activity of the protein of the corresponding naturally-occurring protein yet differ in amino acid sequence due to natural allelic variation or mutagenesis.

The comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm. In an embodiment of the invention, the percent identity between two amino acid sequences is determined using the Needleman and Wunsch ((1970) J. Mol. Biol. 48:444-453) algorithm which has been incorporated into the GAP program in the GCG software package (available at

http://www.gcg.com), using either a Blossum 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. In another embodiment, the percent identity between two nucleotide sequences is determined using the GAP program in the GCG software package (available at http://www.gcg.com), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. In general, percent identity between amino acid sequences referred to herein is determined using the BLAST 2.0 program, which is available to the public at http://www.ncbi.nlm.nih.gov/BLAST. Sequence comparison is performed using an ungapped alignment and using the default parameters (Blossum 62 matrix, gap existence cost of 11, per residue gap cost of 1, and a lambda ratio of 0.85). The mathematical algorithm used in BLAST programs is described in Altschul et al., 1997, *Nucleic Acids Research* 25:3389-3402.

The invention also provides chimeric or fusion proteins. As used herein, a "chimeric protein" or "fusion protein" comprises all or part (e.g., a biologically active portion) of a polypeptide of the invention operably linked to a heterologous polypeptide (i.e., a polypeptide other than the same polypeptide of the invention). Within the fusion protein, the term "operably linked" is intended to indicate that the polypeptide of the invention and the heterologous polypeptide are fused in-frame to each other. The heterologous polypeptide can be fused to the N-terminus or C-terminus of the polypeptide of the invention.

One useful fusion protein is a GST fusion protein in which the polypeptide of the invention is fused to the C-terminus of GST sequences. Such fusion proteins can facilitate the purification of a recombinant polypeptide of the invention.

In another embodiment, the fusion protein contains a heterologous signal sequence at its N-terminus. For example, the native signal sequence of a polypeptide of the invention can be removed and replaced with a signal sequence from another protein. For example, the gp67 secretory sequence of the baculovirus envelope protein can be used as a heterologous signal sequence (*Current Protocols in Molecular Biology*, Ausubel et al., eds., John Wiley & Sons, 1992). Other examples of eukaryotic heterologous signal sequences include the secretory sequences of melittin and human placental alkaline phosphatase (Stratagene; La Jolla, California). In yet another example, useful prokaryotic heterologous signal sequences

include the phoA secretory signal (Sambrook et al., *supra*) and the protein A secretory signal (Pharmacia Biotech; Piscataway, New Jersey).

In yet another embodiment, the fusion protein is an immunoglobulin fusion protein in which all or part of a polypeptide of the invention is fused to sequences derived from a member of the immunoglobulin protein family. The immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject to inhibit an interaction between a ligand (soluble or membrane-bound) and a protein on the surface of a cell (receptor), to thereby suppress signal transduction *in vivo*. The immunoglobulin fusion protein can be used to affect the bioavailability of a cognate ligand of a polypeptide of the invention. Inhibition of ligand/receptor interaction may be useful therapeutically, both for treating proliferative and differentiative disorders and for modulating (e.g., promoting or inhibiting) cell survival. Moreover, the immunoglobulin fusion proteins of the invention can be used as immunogens to produce antibodies directed against a polypeptide of the invention in a subject, to purify ligands and in screening assays to identify molecules which inhibit the interaction of receptors with ligands.

Chimeric and fusion proteins of the invention can be produced by standard recombinant DNA techniques. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and reamplified to generate a chimeric gene sequence (*see, e.g.*, Ausubel et al., *supra*). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g., a GST polypeptide). A nucleic acid encoding a polypeptide of the invention can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the polypeptide of the invention.

A signal sequence of a polypeptide of the invention can be used to facilitate secretion and isolation of the secreted protein or other proteins of interest. Signal sequences are typically characterized by a core of hydrophobic amino acids which are generally cleaved from the mature protein during secretion in one or more cleavage events. Such signal peptides contain processing sites that allow cleavage of the signal sequence from the mature proteins as they pass through the secretory pathway. Thus, the invention pertains to the

described polypeptides having a signal sequence, as well as to the signal sequence itself and to the polypeptide in the absence of the signal sequence (i.e., the cleavage products). In one embodiment, a nucleic acid sequence encoding a signal sequence of the invention can be operably linked in an expression vector to a protein of interest, such as a protein which is ordinarily not secreted or is otherwise difficult to isolate. The signal sequence directs secretion of the protein, such as from a eukaryotic host into which the expression vector is transformed, and the signal sequence is subsequently or concurrently cleaved. The protein can then be readily purified from the extracellular medium by methods known in the art. Alternatively, the signal sequence can be linked to the protein of interest using a sequence which facilitates purification, such as with a GST domain.

The present invention also pertains to variants of the polypeptides of the invention. Such variants have an altered amino acid sequence which can function as either agonists (mimetics) or as antagonists. Variants can be generated by mutagenesis, e.g., discrete point mutation or truncation. An agonist can retain substantially the same, or a subset, of the biological activities of the naturally occurring form of the protein. An antagonist of a protein can inhibit one or more of the activities of the naturally occurring form of the protein by, for example, competitively binding to a downstream or upstream member of a cellular signaling cascade which includes the protein of interest. Thus, specific biological effects can be elicited by treatment with a variant of limited function. Treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein can have fewer side effects in a subject relative to treatment with the naturally occurring form of the protein.

Antibodies

An isolated polypeptide of the invention, or a fragment thereof, can be used as an immunogen to generate antibodies using standard techniques for polyclonal and monoclonal antibody preparation. The full-length polypeptide or protein can be used or, alternatively, the invention provides antigenic peptide fragments for use as immunogens. The antigenic peptide of a protein of the invention comprises at least 8 (e.g., 10, 15, 20, or 30) amino acid residues of the amino acid sequence of a sequence of the invention, e.g., c0148, c0827, and c1083, and encompasses an epitope of the protein such that an antibody raised against the

peptide forms a specific immune complex with the protein. Sequences also useful in the invention include polypeptides encoded by the sequences in Figs. 1, 2A-2R, and 3A-3E or polypeptides encoded by sequences comprising a sequence listed in Figs. 1, 2A-2R, and 3A-3R. Polypeptides encoded by the known genes identified herein as glucose transport-related genes are also useful in the invention.

Epitopes can be encompassed by the antigenic peptide are regions that are located on the surface of the protein, e.g., hydrophilic regions. Hydrophilic regions of selected sequences are indicated in hydrophobicity plots (Figs. 10A-10D, 11A-11D, and 12A-12D). These plots or similar analyses can be used to identify hydrophilic regions in polypeptides useful in the invention.

An immunogen typically is used to prepare antibodies by immunizing a suitable subject, (e.g., rabbit, goat, mouse or other mammal). An appropriate immunogenic preparation can contain, for example, a recombinantly expressed or a chemically synthesized polypeptide. The preparation can further include an adjuvant, such as Freund's complete or incomplete adjuvant, or similar immunostimulatory agent.

Polyclonal antibodies can be prepared as described above by immunizing a suitable subject with a polypeptide of the invention as an immunogen. The antibody titer in the immunized subject can be monitored over time by standard techniques, such as with an enzyme linked immunosorbent assay (ELISA) using immobilized polypeptide. If desired, the antibody molecules can be isolated from the mammal (e.g., from the blood) and further purified by well-known techniques, such as protein A chromatography to obtain the IgG fraction. At an appropriate time after immunization, e.g., when the specific antibody titers are highest, antibody-producing cells can be obtained from the subject and used to prepare monoclonal antibodies by standard techniques, such as the hybridoma technique originally described by Kohler and Milstein, 1975, *Nature* 256:495-497, the human B cell hybridoma technique (Kozbor et al., 1983, *Immunol. Today* 4:72), the EBV-hybridoma technique (Cole et al., 1985, *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96) or trioma techniques. The technology for producing hybridomas is well known (*see generally Current Protocols in Immunology*, 1994, Coligan et al. (eds.) John Wiley & Sons, Inc., New York, NY). Hybridoma cells producing a monoclonal antibody of the invention are detected

by screening the hybridoma culture supernatants for antibodies that bind the polypeptide of interest, e.g., using a standard ELISA assay.

Alternative to preparing monoclonal antibody-secreting hybridomas, a monoclonal antibody directed against a polypeptide of the invention can be identified and isolated by screening a recombinant combinatorial immunoglobulin library (e.g., an antibody phage display library) with the polypeptide of interest. Kits for generating and screening phage display libraries are commercially available (e.g., the Pharmacia *Recombinant Phage Antibody System*, Catalog No. 27-9400-01; and the Stratagene *SurfZAP™ Phage Display Kit*, Catalog No. 240612). Additionally, examples of methods and reagents particularly amenable for use in generating and screening antibody display library can be found in, for example, U.S. Patent No. 5,223,409; PCT Publication No. WO 92/18619; PCT Publication No. WO 91/17271; PCT Publication No. WO 92/20791; PCT Publication No. WO 92/15679; PCT Publication No. WO 93/01288; PCT Publication No. WO 92/01047; PCT Publication No. WO 92/09690; PCT Publication No. WO 90/02809; Fuchs et al., 1991, *Bio/Technology* 9:1370-1372; Hay et al., 1992, *Hum. Antibod. Hybridomas* 3:81-85; Huse et al., 1989, *Science* 246:1275-1281; Griffiths et al., 1993, *EMBO J.* 12:725-734.

Additionally, recombinant antibodies, such as chimeric and humanized monoclonal antibodies, comprising both human and non-human portions, which can be made using standard recombinant DNA techniques, are within the scope of the invention. Such chimeric and humanized monoclonal antibodies can be produced by recombinant DNA techniques known in the art, for example using methods described in PCT Publication No. WO 87/02671; European Patent Application 184,187; European Patent Application 171,496; European Patent Application 173,494; PCT Publication No. WO 86/01533; U.S. Patent No. 4,816,567; European Patent Application 125,023; Better et al., 1988, *Science* 240:1041-1043; Liu et al., 1987, *Proc. Natl. Acad. Sci. USA* 84:3439-3443; Liu et al., 1987, *J. Immunol.* 139:3521-3526; Sun et al., 1987, *Proc. Natl. Acad. Sci. USA* 84:214-218; Nishimura et al., 1987, *Canc. Res.* 47:999-1005; Wood et al., 1985, *Nature* 314:446-449; and Shaw et al., 1988, *J. Natl. Cancer Inst.* 80:1553-1559); Morrison, 1985, *Science* 229:1202-1207; Oi et al., 1986, *Bio/Techniques* 4:214; U.S. Patent 5,225,539; Jones et al., 1986) *Nature* 321:552-525; Verhoeyan et al., 1988, *Science* 239:1534; and Beidler et al., 1988, *J. Immunol.* 141:4053-4060.

Completely human antibodies are particularly desirable for therapeutic treatment of human patients. Such antibodies can be produced using transgenic mice which are incapable of expressing endogenous immunoglobulin heavy and light chains genes, but which can express human heavy and light chain genes. The transgenic mice are immunized in the normal fashion with a selected antigen, e.g., all or a portion of a polypeptide of the invention. Monoclonal antibodies directed against the antigen can be obtained using conventional hybridoma technology. The human immunoglobulin transgenes harbored by the transgenic mice rearrange during B cell differentiation, and subsequently undergo class switching and somatic mutation. Thus, using such a technique, it is possible to produce therapeutically useful IgG, IgA, and IgE antibodies. For an overview of this technology for producing human antibodies, see Lonberg and Huszar (1995, *Int. Rev. Immunol.* 13:65-93). For a detailed discussion of this technology for producing human antibodies and human monoclonal antibodies and protocols for producing such antibodies, *see, e.g.*, U.S. Patent 5,625,126; U.S. Patent 5,633,425; U.S. Patent 5,569,825; U.S. Patent 5,661,016; and U.S. Patent 5,545,806. In addition, companies such as Abgenix, Inc. (Freemont, CA), can be engaged to provide human antibodies directed against a selected antigen using technology similar to that described above.

Completely human antibodies which recognize a selected epitope can be generated using a technique referred to as "guided selection." In this approach a selected non-human monoclonal antibody, e.g., a murine antibody, is used to guide the selection of a completely human antibody recognizing the same epitope. (Jespers et al., 1994, *Biotechnology* 12:899-903).

An antibody directed against a polypeptide of the invention (e.g., monoclonal antibody) can be used to isolate the polypeptide by standard techniques, such as affinity chromatography or immunoprecipitation. Moreover, such an antibody can be used to detect the protein (e.g., in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the polypeptide. The antibodies can also be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, e.g., for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent

materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, beta-galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I, ^{131}I, ^{35}S or ^{3}H.

Screening Assays

The invention provides a method for identifying modulators, i.e., candidate agents or reagents, of expression or activity of a glucose transport-related nucleic acid or polypeptide. Such candidate agents or reagents include polypeptides, oligonucleotides, peptidomimetics, carbohydrates or small molecules such as small organic or inorganic molecules (e.g., non-nucleic acid small organic chemical compounds) that modulate expression (protein or mRNA) or activity of one or more glucose transport-related polypeptides or nucleic acids. In general, screening assays involve assaying the effect of a test agent on expression or activity of a glucose transport-related nucleic acid or polypeptide in a test sample (i.e., a sample containing the glucose transport-related nucleic acid or polypeptide). Expression or activity in the presence of the test compound or agent is compared to expression or activity in a control sample (i.e., a sample containing a glucose transport-related polypeptide that was not incubated in the presence of the test compound). A change in the expression or activity of the glucose transport-related nucleic acid or polypeptide in the test sample compared to the control indicates that the test agent or compound modulates expression or activity of the glucose transport-related nucleic acid or polypeptide and is a candidate agent.

In one embodiment, the invention provides assays for screening candidate agents that bind to or modulate the activity of a polypeptide or nucleic acid of the invention or biologically active portion thereof. The compounds to be screened, can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the "one-bead one-compound"

library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds (Lam, 1997, *Anticancer Drug Des.* 12:145).

Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt et al., 1993, *Proc. Natl. Acad. Sci. USA* 90:6909; Erb et al., 1994, *Proc. Natl. Acad. Sci. USA* 91:11422; Zuckermann et al., 1994, *J. Med. Chem.* 37:2678; Cho et al., 1993, *Science* 261:1303; Carrell et al., 1994, *Angew. Chem. Int. Ed. Engl.* 33:2059; Carell et al., 1994, *Angew. Chem. Int. Ed. Engl.* 33:2061; and Gallop et al., 1994, *J. Med. Chem.* 37:1233.

Libraries of compounds may be presented in solution (e.g., Houghten, 1992, *Bio/Techniques* 13:412-421), or on beads (Lam, 1991, *Nature* 354:82-84), chips (Fodor, 1993, *Nature* 364:555-556), bacteria (U.S. Patent No. 5,223,409), spores (Patent Nos. 5,571,698; 5,403,484; and 5,223,409), plasmids (Cull et al., 1992, *Proc. Natl. Acad. Sci. USA* 89:1865-1869) or phage (Scott and Smith, 1990, *Science* 249:386-390; Devlin, 1990, *Science* 249:404-406; Cwirla et al., 1990, *Proc. Natl. Acad. Sci. USA* 87:6378-6382; and Felici, 1991, *J. Mol. Biol.* 222:301-310).

In one embodiment, the assay is a cell-based assay in which a cell expressing a polypeptide of the invention, or a biologically active portion thereof, on the cell surface is contacted with a test compound. The ability of the test compound to bind to the polypeptide is then determined. The cell, for example, can be a yeast cell or a cell of mammalian origin. Determining the ability of the test compound to bind to the polypeptide can be accomplished, for example, by coupling the test compound with a radioisotope or enzymatic label such that binding of the test compound to the polypeptide or biologically active portion thereof can be determined by detecting the labeled compound in a complex. For example, test compounds can be labeled with ^{125}I, ^{35}S, ^{14}C, or ^{3}H, either directly or indirectly, and the radioisotope detected by direct counting of radioemmission or by scintillation counting. Alternatively, test compounds can be enzymatically labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product. In one embodiment, the assay comprises contacting a cell which expresses a membrane-bound form of a polypeptide of the invention,

or a biologically active portion thereof, on the cell surface with a known compound which binds to the polypeptide to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with the polypeptide, wherein determining the ability of the test compound to interact with the polypeptide comprises determining the ability of the test compound to preferentially bind to the polypeptide or a biologically active portion thereof as compared to the known compound.

In another embodiment, an assay is a cell-based assay comprising contacting a cell expressing a membrane-bound form of a polypeptide of the invention, or a biologically active portion thereof, on the cell surface with a test compound and determining the ability of the test compound to modulate (e.g., stimulate or inhibit) the activity of the polypeptide or biologically active portion thereof. Determining the ability of the test compound to modulate the activity of the polypeptide or a biologically active portion thereof can be accomplished, for example, by determining the ability of the polypeptide to bind to or interact with a target molecule.

Determining the ability of a polypeptide or nucleic acid of the invention to bind to or interact with a target molecule can be accomplished by one of the methods described herein for determining direct binding. As used herein, a "target molecule" is a molecule with which a selected polypeptide or nucleic acid (e.g., a polypeptide or nucleic acid of the invention) binds or interacts with in nature, for example, a molecule on the surface of a cell which expresses the selected protein, a molecule on the surface of a second cell, a molecule in the extracellular milieu, a molecule associated with the internal surface of a cell membrane or a cytoplasmic molecule. A target molecule can be a polypeptide or nucleic acid of the invention or some other polypeptide, protein or nucleic acid. For example, a target molecule can be a component of a signal transduction pathway which facilitates transduction of an extracellular signal (e.g., a signal generated by binding of a compound to a polypeptide of the invention) through the cell membrane and into the cell or a second intercellular protein which has catalytic activity or a protein which facilitates the association of downstream signaling molecules with a polypeptide of the invention. Determining the ability of a polypeptide of the invention to bind to or interact with a target molecule can also be accomplished by determining the activity of the target molecule. For example, the activity of the target molecule can be determined by detecting induction of a cellular second messenger of the

target (e.g., intracellular Ca^{2+}, diacylglycerol, or IP3), detecting catalytic/enzymatic activity of the target on an appropriate substrate, detecting the induction of a reporter gene (e.g., a regulatory element that is responsive to a polypeptide of the invention operably linked to a nucleic acid encoding a detectable marker, e.g., luciferase), or detecting a cellular response, for example, cellular differentiation, or cell proliferation. When the target molecule is a nucleic acid, the compound can be, e.g., a ribozyme or antisense molecule.

In yet another embodiment, an assay of the present invention is a cell-free assay comprising contacting a polypeptide or nucleic acid of the invention, or biologically active portion thereof, with a test compound and determining the ability of the test compound to bind to the polypeptide or biologically active portion thereof. Binding of the test compound to the polypeptide can be determined either directly or indirectly as described above. In one embodiment, the assay includes contacting the polypeptide of the invention or biologically active portion thereof with a known compound which binds the polypeptide to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with the polypeptide (e.g., its ability to compete with binding of the known compound), wherein determining the ability of the test compound to interact with the polypeptide comprises determining the ability of the test compound to preferentially bind to the polypeptide or biologically active portion thereof as compared to the known compound. When the test compound is targeted to a nucleic acid, the binding of the test compound to the nucleic acid can be tested, e.g., by binding, by fragmentation of the nucleic acid (as when the test compound is a ribozyme), or by inhibition of transcription or translation in the presence of the test compound.

In another embodiment, an assay is a cell-free assay comprising contacting a polypeptide of the invention or biologically active portion thereof with a test compound and determining the ability of the test compound to modulate (e.g., stimulate or inhibit) the activity of the polypeptide or biologically active portion thereof. For example, determining the ability of the test compound to modulate the activity of the polypeptide can be accomplished by determining the ability of the polypeptide of the invention to modify the target molecule. Such methods can, alternatively, measure the catalytic/enzymatic activity of the target molecule on an appropriate substrate. In general, modulation of the activity of the polypeptide of the invention or biologically portion thereof is determined by comparing the

activity in the absence of the test compound to the activity in the presence of the test compound.

In yet another embodiment, the cell-free assay comprises contacting a polypeptide or nucleic acid of the invention, or biologically active portion thereof, with a known compound which binds to the polypeptide to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with the polypeptide or nucleic acid, wherein determining the ability of the test compound to interact with the polypeptide or nucleic acid comprises determining the ability of the polypeptide or nucleic acid to preferentially bind to or modulate the activity of a target molecule.

The cell-free assays of the present invention are amenable to use of either a soluble form or the membrane-bound form of a polypeptide of the invention. In the case of cell-free assays comprising the membrane-bound form of the polypeptide, it may be desirable to utilize a solubilizing agent such that the membrane-bound form of the polypeptide is maintained in solution. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-octylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton X-100, Triton X-114, Thesit, Isotridecypoly(ethylene glycol ether)n, 3-[(3-cholamidopropyl)dimethylamminio]-1-propane sulfonate (CHAPS), 3-[(3-cholamidopropyl)dimethylamminio]-2-hydroxy-1-propane sulfonate (CHAPSO), or N-dodecyl-N,N-dimethyl-3-ammonio-1-propane sulfonate.

In more than one embodiment of the above assay methods of the present invention, it may be desirable to immobilize either the polypeptide of the invention or its target molecule to facilitate separation of complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to the polypeptide, or interaction of the polypeptide with a target molecule in the presence and absence of a test agent, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtitre plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided which adds a domain that allows one or both of the proteins to be bound to a matrix. For example, glutathione-S-transferase fusion proteins or glutathione-S-transferase fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical; St. Louis, MO) or glutathione derivatized microtitre plates, which are then combined with the test compound or the test compound and

either the non-adsorbed target protein or a polypeptide of the invention, and the mixture incubated under conditions conducive to complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads or microtitre plate wells are washed to remove any unbound components and complex formation is measured either directly or indirectly, for example, as described above. Alternatively, the complexes can be dissociated from the matrix, and the level of binding or activity of the polypeptide of the invention can be determined using standard techniques.

Other techniques for immobilizing proteins on matrices can also be used in the screening assays of the invention. For example, either the polypeptide of the invention or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated polypeptide of the invention or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques well known in the art (e.g., biotinylation kit, Pierce Chemicals; Rockford, IL), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with the polypeptide of the invention or target molecules but which do not interfere with binding of the polypeptide of the invention to its target molecule can be derivatized to the wells of the plate, and unbound target or polypeptide of the invention trapped in the wells by antibody conjugation. Methods for detecting such complexes such as GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the polypeptide of the invention or target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the polypeptide of the invention or target molecule.

In another embodiment, modulators of expression of a polypeptide of the invention are identified in a method in which a cell is contacted with a test agent or compound and the expression of the selected mRNA or protein (i.e., the mRNA or protein corresponding to a polypeptide or nucleic acid of the invention) in the cell is determined. The level of expression of the selected mRNA or protein in the presence of the test agent is compared to the level of expression of the selected mRNA or protein in the absence of the test agent. The test agent can then be identified as a modulator of expression of the polypeptide (i.e., a candidate compound)of the invention based on this comparison. For example, when expression of the selected mRNA or protein is greater (statistically significantly greater) in the presence of the test agent than in its absence, the test agent is identified as a candidate

agent that is a stimulator of the selected mRNA or protein expression. Alternatively, when expression of the selected mRNA or protein is less (statistically significantly less) in the presence of the test agent than in its absence, the test agent is identified as a candidate agent that is an inhibitor of the selected mRNA or protein expression. The level of the selected mRNA or protein expression in the cells can be determined by methods described herein.

In yet another aspect of the invention, a polypeptide of the inventions can be used as "bait proteins" in a two-hybrid assay or three hybrid assay (*see, e.g.*, U.S. Patent No. 5,283,317; Zervos et al., 1993, *Cell* 72:223-232; Madura et al., 1993, *J. Biol. Chem.* 268:12046-12054; Bartel et al., 1993, *Bio/Techniques* 14:920-924; Iwabuchi et al., 1993, *Oncogene* 8:1693-1696; and PCT Publication No. WO 94/10300), to identify other proteins, that bind to or interact with the polypeptide of the invention and modulate activity of the polypeptide of the invention. Such binding proteins are also likely to be involved in the propagation of signals by the polypeptide of the inventions as, for example, upstream or downstream elements of a signaling pathway involving the polypeptide of the invention.

Electronic Data Storage and Processing

The invention includes nucleic acid and polypeptide sequences that are provided in digital form that can be transmitted and read electronically (e.g., in a database). In some embodiments, the database can be queried for comparison with data provided (e.g., a nucleic acid sequence or a pattern of expression). All sequence information or data provided for comparison with the database can be transmitted to the database, e.g., by email, via the Internet, on diskette, or any other mode of electronic or non-electronic communication.

The invention thus features an electronic method of determining whether a patient has a glucose-transport-related disorder by obtaining an electronic form of a nucleic acid sequence from the patient; obtaining a database of nucleic acid molecules whose expression is altered in a glucose transport-related disorder such as type II diabetes that includes nucleic acid molecules of individuals with glucose-transport related disorders; and comparing the patient nucleic acid sequence with the nucleic acid molecules in the database, wherein a patient nucleic acid sequence that matches a nucleic acid molecule in the database indicates the patient has or is at risk for a glucose-transport related disorder.

The invention also includes a database that includes an electronic form (e.g., digital form) of the nucleic acid molecules of the invention, and a computer-readable instructions for a processor to carry out the comparison method. The database can also be stored on a machine- or computer-readable medium, and can be accessed, e.g., through a communications network, such as the Internet.

As used herein, "sequence information" refers to any nucleotide and/or amino acid sequence information, including but not limited to full-length nucleotide and/or amino acid sequences, partial nucleotide and/or amino acid sequences. Moreover, information "related to" the sequence information includes detecting the presence or absence of a sequence (e.g., detection of expression of a sequence, fragment, or polymorphism), determination of the level of a sequence (e.g., detection of a level of expression, for example, a quantitative detection), detection of a reactivity to a sequence (e.g., detection of protein expression and/or levels, for example, using a sequence-specific antibody), detection of a pattern of expression of two or more sequences, and the like. These sequences can be read by electronic apparatus and can be stored on any suitable medium for storing, holding, or containing data or information that can be read and accessed by an electronic apparatus. Such media can include, but are not limited to: magnetic storage media, such as floppy disks, hard disk storage medium, and magnetic tape; optical storage media such as compact disks; electronic storage media such as RAM, ROM, EPROM, EEPROM and the like; general hard disks and hybrids of these categories such as magnetic/optical storage media. The medium is adapted or configured for having recorded thereon sequence information.

As used herein, the term "electronic apparatus" is intended to include any suitable computing or processing apparatus or other device configured or adapted for storing data or information. Examples of electronic apparatus suitable for use with the present invention include stand-alone computing apparatus such as personal computers (PCs) and large computer systems. These systems can be accessed by communications networks, including local area networks (LAN), wide area networks (WAN), Internet, Intranet, and Extranet. For example, the database can be made available on an Internet website.

As used herein, "stored" refers to a process for encoding information on the electronic apparatus readable medium. Those skilled in the art can readily adopt any of the presently

known methods for recording information on known media to generate manufactures comprising the sequence information.

A variety of software programs and formats can be used to store the sequence information on the electronic apparatus readable medium. For example, the sequence information can be represented in a word processing text file, formatted in commercially-available software such as WordPerfect® and MicroSoft® Word®, or represented in the form of an ASCII file, stored in a database application, such as DB2®, Sybase® , Oracle®, or the like, as well as in other forms. Any number of data processor structuring formats (e.g., text file or database) can be employed to obtain or create a medium having recorded thereon the sequence information.

By providing sequence information in machine or computer-readable form, one can routinely access the sequence information for a variety of purposes. For example, one skilled in the art can use the sequence information in computer-readable form to compare a specific sequence with the sequence information stored within a database. Search means are used to identify fragments or regions of the sequences that match a particular sequence.

The present invention therefore provides a medium for storing or holding a database or instructions for performing a method for determining whether an individual has a specific disease or disorder related to glucose transport or a pre-disposition for a specific disease or disorder related to glucose transport, wherein the method can include analyzing the individual's sequence information and based on the sequence information, determining whether the individual has a particular disorder or a predisposition for a particular disorder associated with a specific genetic sequence, and/or recommending a particular treatment for the disorder or pre-disorder condition. For example, the pattern of expression of glucose transport-related sequences or proteins from an individual suspected of having a glucose transport-related disorder (e.g., type II diabetes) can be analyzed, and, based on the analysis (e.g., aberrant expression of one or more glucose transport-related genes), a diagnosis provided and instructions for treatment.

The invention will be further described in the following examples which do not limit the scope of the invention described in the claims.

EXAMPLES

Three approaches were used to identify genes and proteins involved in glucose transport. First, several subtractive cDNA libraries were constructed that consist of genes selectively expressed in insulin-responsive tissues. Furthermore, it has been discovered that at least two of these genes have a role in regulating GLUT4 translocation. As a second approach, microarrays were screened with fluorescently labeled probes synthesized from mRNA isolated from insulin-responsive tissues. In the third approach, a subcellular fraction was prepared that was enriched for vesicles involved in glucose transport. Proteins from this fraction were prepared and analyzed using microsequencing techniques. Additional analysis comparing the predicted protein sequences obtained in the first two approaches with the vesicle protein sequences provided a subset of sequences involved in glucose transport that are useful for certain aspects of the invention.

Example 1: Subtractive Libraries

Two methods were used to construct subtractive libraries.

In the first method, suppression subtractive hybridization was used (Diatchenko et al., 1996, Proc Natl. Acad. Sci U S A 93:6025-30). In this method, a first library was constructed that consisted of sequences that are highly expressed in muscle, but not in 3T3-L1 fibroblasts (available from American Type Culture Collection; ATCC). The second library consisted of sequences that are highly expressed in 3T3-L1 adipocytes, but not in 3T3-L1 fibroblasts. The general method for this procedure is diagrammed in Fig. 4.

Libraries were constructed by reverse transcription of total mRNA isolated from plates of confluent 3T3-L1 fibroblasts and 3T3-L1 adipocytes 9 to 10 days after the start of differentiation. The resulting cDNAs were then digested with the restriction enzyme Rsa I. Digested adipocyte cDNA was divided into two pools, and each pool was ligated to a different oligonucleotide adaptor. Adaptor 1 was:

5'-CTAATACGACTCACTATAGGGCTCGAGCGGCCGCCCGGGCAGGT-3' (SEQ ID NO:94)
GGCCCGTCCA-5' (SEQ ID NO:95)

Adaptor 2 was:

5'-CTAATACGACTCACTATAGGGCAGCGTGGTCGCGGCCGAGGT-3' (SEQ ID NO:96)
GCCGGCTCCA-5' (SEQ ID NO:97)

Each pool of adipocyte cDNA (tester DNA) was then hybridized with an excess of fibroblast cDNA (driver DNA) for 9 hours at 68° C. The two hybridization mixtures were combined and incubated overnight at 68° C. After hybridization the 5' overhangs were filled in with Taq DNA polymerase, and amplified by PCR using primers that are homologous to each of the adaptors. This subtraction procedure was also performed using the mouse muscle cDNA as the tester, and 3T3-L1 fibroblast cDNA as the driver.

As a test to demonstrate that muscle and adipocyte specific transcripts are amplified by this procedure, the final products of both subtractions were amplified using PCR primers internal to GLUT4 and α-tubulin transcripts. The final product of muscle subtraction (SUB) and the unsubtracted muscle cDNA (UNSUB) were used for PCR analysis with primers internal to the coding regions of Glut4 (G4) and α-tubulin-1 (TUB). Glut4 and α-tubulin-1 primers were designed to amplify 485 bp and 408 bp fragments respectively. PCR samples were removed after 23, 28 and 33 PCR cycles and loaded onto a 1.5% TAE (40mM Tris-acetate, pH8.0 1mM EDTA) agarose gel. The gel was stained with ethidium bromide and visualized with UV light. As expected, GLUT4 cDNA (representing GLUT4 expression) was found in the subtracted muscle cDNA but tubulin cDNA was present in relatively small amounts because tubulin is expressed in both fibroblasts and muscle (and so a substantial amount of the tubulin sequence was subtracted out). GLUT4 is expressed in muscle but not in fibroblasts, and so, as expected present in relatively large amounts. In the muscle-subtracted cDNA, the GLUT4 signal is stronger in earlier PCR cycles, while the tubulin signal in suppressed. Similar results were obtained with PCR analysis with 3T3-L1 adipocyte-subtracted cDNA.

To construct the libraries, the final PCR products from the 3T3-L1 adipocyte subtraction were digested with Rsa I and cloned into Eco RV restricted the pBluescript SK+ vector (STRATAGENE®) creating a library of adipocyte subtractive clones. The library contained approximately 2×10^3 clones. The cloned plasmid DNA sequences were analyzed by dideoxy sequencing with either the M13 -20 or reverse primer on an ABI 377 automatic sequencer. In an initial round of sequencing, 183 independent clones, representing expression from 65 different genes, were sequenced. Sequences were analyzed in a search

against the non-redundant (NR) nucleotide database using the Blast program at www.ncbi.nlm.nih.gov/blast/blast.cgi. The gapped BLAST program was used against the non-redundant or the dbest database. All BLAST searches were performed using the default settings which are: Expect=10; Filter for Low complexity: on; Filter for Human Repeats: off; Mask for lookup table only: off; Matrix=Blosum62; Gap existence cost=11; Per residue gap cost=1; Lambda ratio=85.

Genes previously shown to be preferentially expressed or not preferentially expressed in adipocytes are those in which their mRNA expression profiles have been published in journal articles in the Medline database. A summary of these sequences is shown in Figs. 6A-6E. Approximately 60% of the sequenced clones in this library were from genes previously reported as overexpressed in 3T3-L1 adipocytes. Another 23% of the clones consisted of known gene sequences whose expression pattern was known in adipocytes, while 13% of the sequenced clones had unknown (previously unreported) sequences. Four percent of the cloned sequences are from genes of mitochondrial origin. The identity of the genes in the subtractive library that have already shown to be preferentially expressed in 3T3-L1 adipocytes are listed in Figs. 6A-6E. Genes, such as adipoQ and stearoyl-CoA desaturase, that are found at the highest frequency in this subtractive library are also those that were discovered in previous attempts to clone genes that are highly expressed in 3T3-L1 adipocytes upon differentiation (Ntambi et al., 1988, J. Biol. Chem 263:17291-17300; Bernlohr et al., 1984, Proc. Nat. Acad. Sci. USA 81: 5468-5472; Hu et al., 1996, J. Biol. Chem. 271:10697-10703; Min and Spiegelman, 1986, Nucleic Acids Res. 14:8879-8892).

Sequences that are expressed in the Adipocyte Subtractive library that are from genes with unknown function are listed in Figs. 3A-3E.

Example 2: Construction of a Muscle-Adipocyte Library

To identify genes encoding proteins that are involved in glucose transport, gene expression in 3T3-L1 adipocytes and muscle was investigated. To accomplish this, another library was constructed consisting of genes that fulfilled the following two criteria. First, the genes had to be highly expressed in both 3T3-L1 adipocytes and mouse skeletal muscle; second, the genes could not be highly expressed in 3T3-L1 fibroblasts. This library, the Muscle-Adipocyte Union Library (MAU library), was constructed using a modification of

the suppression subtractive hybridization technique (Fig. 5). The method was like the subtractive suppression modification technique described in Fig. 4 except that adaptor 1 was ligated to Rsa I-digested 3T3-L1 adipocyte cDNA while adaptor 2 was ligated to Rsa I-digested mouse muscle cDNA. Both cDNAs were then hybridized to an excess of 3T3-L1 fibroblast DNA. The two hybridization reactions were then mixed to create hybrid molecules in which one strand originated from adipocytes and the second strand of the hybrid was from muscle. Because only these hybrid molecules have different adaptors on each end, they can be PCR amplified, unlike the rest of the cDNAs. These hybrid products were then amplified using PCR. The final PCR products of the 3T3-L1 muscle-adipocyte union subtraction were cloned into overhang vector pCR2.1. (INVITROGEN®) to produce a library of approximately 10^4 clones. Plasmid DNAs were dideoxy sequenced with the either the M13-20 or reverse primer on an ABI 377 automatic sequencer. Sequences were searched against the non-redundant (NR) nucleotide database using the Blast program at www.ncbi.nlm.nih.gov/blast/blast.cgi. Genes previously shown to be overexpressed or not overexpressed in adipocytes are those in which their mRNA expression profiles has been published in journal articles in the Medline database. Figs. 7A-7U show the summary of sequences from this library. These clones represent as many as 265 different genes. About 40% of these sequences are expressed from genes that have previously been shown to be preferentially in muscle, adipocytes, or both tissues. Another 26% of the clones are sequences from known genes whose expression profile is not known, and 17% of the clones represent previously unidentified genes. A large percentage of sequences (12%) represent genes of mitochondrial origin. Fig. 1 shows sequences from this library that are novel, and Figs. 2A-2R show the sequences of selected clones from this library. Figs. 7A-7U show the genes that encode the sequences identified in the MAU library including the GenBank accession no., when one is known. Figs. 7A-7U also list the homologous human genes for these sequences and the expression profile of each sequence with respect to its expression in adipocytes and muscle.

The sequences identified in this manner are useful, e.g., for detecting a glucose transport-related disorder such as type II diabetes.

Example 3: mRNA Expression Profiles of Unknown Genes in the 3T3-L1 Adipocyte Subtractive and the Muscle-Adipocyte Union Libraries.

To determine the expression profile of cognate RNA from library clones that have not been previously reported to be overexpressed (i.e., preferentially expressed) in insulin-sensitive (e.g., adipocyte and muscle) tissues, expression of these sequences was analyzed in undifferentiated 3T3-L1 cells and differentiated 3T3-L1 adipocytes. Northern blot analysis was used in which 3T3-L1 and mouse multi-tissue Northern blots were probed.

Cloned inserts from the Adipocyte Subtractive Library clones were labeled with ^{32}P-dCTP and used in an initial screen to probe Northern blots of total RNA from 3T3-L1 fibroblasts and adipocytes. For Northern blotting, 3T3-L1 and multiple tissue total RNA (10 μg) were electrophoresed on 1.2% agarose/6.6% formaldehyde gels, then transferred to Nytran membranes. Before transfer, gels were stained with ethidium bromide and visualized with UV light in order to confirm equal loading of RNAs. Blots were probed with inserts containing fragments of previously unidentified genes from both libraries Probes were labeled with P^{32}-dCTP and incubated with the membranes overnight at 42°C. Blots were washed twice with 2x SSC/0.1%SDS at room temperature, twice in 0.2x SSC/0.1% SDS at room temperature and twice in 0.2x SSC/0.1% SDS at 42°C. After washing, blots were exposed to a phosphor screen for one to three days. Phosphor screens were scanned with the Storm 860 Scanner from Molecular Dynamics. Full-length clones for many of these unknown genes have been obtained either by purchasing IMAGE Consortium clones or by screening muscle or adipocyte lambda libraries (such libraries can be made using methods known in the art).

Seventy-eight clones from the Adipocyte Subtractive Library were characterized. Sixty of the 78 cloned sequences (approximately 75%) were preferentially expressed upon adipocyte differentiation (i.e., in 3T3-L1 adipocytes).

Thirty-two clones from the 3T3-L1 Muscle-Adipocyte Union library (MAU library) were analyzed. Nineteen were preferentially expressed in 3T3-L1 adipocytes. This leads to the conclusion that approximately 50% of the clones in the MAU library, whose expression has not previously been reported, are preferentially expressed in 3T3-L1 adipocytes. This indicates that approximately 80% of the clones in the 3T3-L1 Adipocyte Subtractive Library and 70% of the clones in the Muscle-Adipocyte Union Library (MAU library) are highly

expressed in at least one insulin-sensitive tissue. (For the 3T3-L1 Adipocyte Subtractive Library, 60% of sequences previously shown to be preferentially expressed + ½ of 40% = 80%; for MAU library, 40% of sequences previously shown to be preferentially expressed + ½ of 60% of uncharacterized genes = 70%). Genes that were found to be preferentially expressed in 3T3-L1 adipocytes were used to probe mouse multi-tissue Northern blots. Using Northern analysis, it was confirmed that 11 previously unidentified genes from the MAU library (i.e., genes expressed in adipocytes and muscle) are expressed in at least two different insulin-sensitive tissues (see Figs. 6A-6E and 7A-7U; "overexpressed" indicates that the sequence was found to be preferentially expressed in insulin-sensitive cells in these experiments).

Using multi-tissue Northern blots it was shown that that six previously identified genes are highly expressed in insulin-sensitive tissues. Furthermore, at least two of these proteins have a role in regulating GLUT4. This was determined as follows. Three clones in the Muscle-Adipocyte Union Library consist of the 3' end of PP2Cα1 (Genbank Acession No.D28117 Kato et al., 1994, Gene 145:311-312). Northern blot analysis demonstrated that at least three transcripts of PP2Cα are highly expressed in both 3T3-L1 adipocytes and in mouse fat. We further examined mRNA expression of PP2Cα. For Northern blotting, 3T3-L1 and multiple tissue total RNA (10 μg) were separated by electrophoresis on 1.2% agarose/ 6.6% formaldehyde gels, then transferred on to Nytran membranes. Blots were probed with library clone c0452, which contains the last 216 base pairs of the PP2Cα1 coding sequence along with the 288 base pairs of 3' noncoding region. Probes were labeled with P^{32}-dCTP and incubated with the membranes overnight at 42°C. Blots were washed twice with 2x SSC/0.1%SDS at room temperature, twice in 0.2x SSC/0.1% SDS at room temperature and twice in 0.2x SSC/0.1% SDS at 42°C. After washing, 3T3-L1 blots were exposed to film for one day, while multi tissue northern blots were exposed to a phosphor screen for one to three days. Phosphor screens were scanned with the Storm 860 Scanner from Molecular Dynamics.

To assess the role of PP2Cα1 in insulin-stimulated glucose transport, PP2Cα1 protein was microinjected into 3T3-L1 adipocytes, and GLUT4 translocation was determined by immunofluorescence. Microinjection of PP2Cα1 was found to potentiate the ability of a submaximal 1 nM concentration of insulin to translocate GLUT4 to the plasma membrane to

levels close if not equal to that of a maximal 10nM insulin stimulation. To examine the effect of microinjected PP2Cα1 on GLUT4 translocation, 3T3-L1 adipocytes were incubated in serum free medium for two hours and microinjected with either IgG alone or PP2Cα along with IgG. Sixty minutes later adipocytes were incubated with media alone, 1 nM insulin or a maximally effective concentration of insulin (10 nM) for 30 minutes. Cells were then fixed with methanol and then stained with anti-GLUT4 antibody. Adipocytes were examined using fluorescence microscopy (Zeiss Axioskop, at 630x magnification) and scored for scored for the presence of substantial cell surface GLUT4 immunoreactivity at the plasma membrane. Controls are cells on the same coverslips that were not injected. Microinjection of phosphatases 2A or 2B had no effect on the ability of insulin to activate GLUT4 translocation. Western blotting has also revealed that PP2Cα selectively co-immunoprecipitates insulin receptors but not PDGF receptors in an insulin-enhanced manner.

Gα11 (Q209L) Induced 2-Deoxyglucose Uptake in Differentiated 3T3-L1 Adipocytes.

Gα11 sequence (Genbank Accession No. U37411; Davignon et al., 1996, Genomics 31:359-366) was identified in the 3T3-L1 Adipocyte Subtractive Library. This protein is a member of the Gαq family which are heterotimeric components of G protein complexes. Northern blot analysis confirmed that Gα11 expression is induced upon 3T3-L1 adipocyte differentiation, and that it is more abundant by far in fat than in any other tissue. Differentiated 3T3-L1 adipocytes were seeded at 150,000 cells per well in 24 well plates and then infected with either control or Gα11 (Q209L) adenoviruses. Thirty hours after infection, plates were serum starved for two hours in Krebs-Ringer phosphate buffer with BSA and pyruvate. Plates were then treated with or without wortmannin (a specific inhibitor of PI3 kinase) for 15 minutes followed by stimulation with insulin or endothelin for 30 minutes. Cells were then assayed for 2-deoxyglucose uptake as described in Frost and Lane (1985, J. Biol. Chem. 260:2646-2652).

For Northern blotting, 3T3-L1 and multiple tissue total RNA (10 μg) were separated on 1.2% agarose/ 6.6% formaldehyde gels, then transferred on to Nytran membranes. Blots were probed with library clone b0031, which contains nt 237 to nt 435 of the Gα11 coding sequence. Probes were labeled with P^{32}-dCTP and incubated with the membranes overnight at 42°C. Blots were washed twice with 2x SSC/0.1%SDS at room temperature, twice in 0.2x

SSC/0.1% SDS at room temperature and twice in 0.2x SSC/0.1% SDS at 42°C. After washing, 3T3-L1 blots were exposed to film for one day, while multi tissue northern blots were exposed to a phosphor screen for three days. Phosphor screens were scanned with the Storm 860 Scanner from Molecular Dynamics. A closely related protein Gq did not have this expression profile. Infection of 3T3-L1 adipocytes with a recombinant adenovirus expressing a constitutively active form of Gα11 expression, but not the native protein led to an increase in GLUT4 concentration in the plasma membrane, and a fourfold increase in glucose uptake in a wortmannin-insensitive manner. Thus, wortmannin does not inhibit the ability of the active form of Gα11 to stimulate GLUT4 translocation.

Since PI3 kinase activation is required for insulin to activate GLUT4 translocation, these data indicate that Gα11 is likely a mediator of PI3 kinase independent activators of GLUT4 translocation, such as endothelin. In addition, these data demonstrate that glucose transport-related genes were identified using the methods described herein. They also illustrate an assay for identifying glucose transport-related sequences that are PI3 kinase independent activators of GLUT4 translocation.

Example 4: Polypeptides Isolated from GLUT4-Enriched Vesicles

The GLUT4 glucose transporter resides primarily in perinuclear membranes in unstimulated 3T3-L1 adipocytes and is acutely translocated to the cell surface in response to insulin. A novel method of purifying intracellular GLUT4-enriched membranes was used to identify polypeptides involved in glucose transport.

Antibodies

Rabbit polyclonal anti-GLUT4 antibody was raised against the C-terminal 12 amino acid sequence of GLUT4. Mouse anti-transferrin receptor was from Zymed. Rabbit polyclonal anti-VAMP2 antibody was from StressGen Biotechnologies Corp. Mouse monoclonal anti-vimentin antibody used in immunoblots and immuno-electron microscopy analysis was from Santa Cruz. Mouse monoclonal anti-α-tubulin antibody, used in immunoblot and immuno-electron microscopy analysis and the secondary antibodies conjugated to gold particles for immuno-electron microscopy were from Amersham Pharmacia Biotech.

Immunoblotting

Fractions from velocity gradients and equilibrium density gradient were prepared as described above and aliquots from these fractions were subjected to SDS-PAGE on resolving gels according to Laemmli (1970, Nature 227:680-685). Separated proteins were electrophoretically transferred to nitrocellulose membrane, blocked with 3% nonfat milk and 1% BSA in TTBS (0.05% Tween 20 in Tris-buffered saline) and then incubated with primary antibody in TTBS containing 1% BSA. After incubation, membranes were washed with TTBS and incubated with horseradish peroxidase-labeled anti-mouse IgG for the detection of monoclonal antibodies or with horseradish peroxidase-labeled anti-rabbit IgG for detection of polyclonal antibodies. Proteins were visualized using an enhanced chemiluminescent substrate kit (Amersham Pharmacia Biotech) and immunoblot intensities were quantified by a scanning densitometer.

Electron Microscopy

GLUT4-containing membranes of the insulin sensitive fractions from the equilibrium density gradient were isolated as described above. Fractions were pooled, pelleted by centrifugation at 48,000 rpm for 2 hours, resuspended in PBS and fixed in a final concentration of 2% paraformaldehyde in PBS. GLUT4-vesicles were then adsorbed to Formvard-coated gold grids and processed for double labeling as outlined in Martin et al. (*supra*) and Sleeman et al. (1998, J. Biol. Chem. 273:3132-3135). Grids were incubated with 50μl of primary antibody diluted in 1% BSA and PBS as follows: anti-GLUT4, anti-IRAP, anti-vimentin, anti-α-tubulin or non-immune IgG, as a negative control. After incubation with each IgG fraction, grids were labeled with either 5 or 15 nm gold particles conjugated to the secondary antibody (goat anti-rabbit or goat anti-mouse). Grids were stained with 1% uranyl acetate, dried and viewed using a transmission electron microscope PHILLIPS CM.10.

Purification of insulin-responsive GLUT4-containing membranes

GLUT4-containing membranes were prepared by first isolating low density (LD) microsomes then subjecting these to further purification on sucrose velocity gradients.

Finally, the GLUT4 fractions from the sucrose gradients were subjected to equilibrium density sucrose gradients. The preparations were made from primary, unstimulated or insulin stimulated rat adipocytes, although the could also be prepared from other tissues, e.g., striatal muscle.

To prepare the initial crude membrane preparations for purification, adipocytes were isolated from epididymal fat pads of Male Sprague-Dawley Rats (125-150 g) by collagenase digestion in Krebs-Ringer/HEPES, pH 7.4, supplemented with 2% bovine serum albumin and 2 mM pyruvate. Following digestion, the cells were washed and permitted to recover for 30 minutes. The cells were then incubated at 37°C with or without 100 nM insulin for 20 minutes. The cells were washed with PBS and immediately homogenized in buffer A (50 mM HEPES, pH 7.4, 10 mM NaF, 1 mM NaPPi, 0.1 mM Na_3VO_4, 1 mM phenylmethylsulfonyl fluoride, 10 μg/ml aprotinin, and 10 μg/ml leupeptin), and then subjected to differential centrifugation as described in Czech and Buxton, 1993, J. Biol. Chem. 268:9187-9190. Low density microsomes were prepared by modifications of previously described methods (Mackeell, D.W. and Jarret, L., 1970, J. Cell Biol., 44:417432). Briefly, cells were homogenized for 15 strokes with a motor-driven Teflon/glass homogenizer in 24 ml of buffer containing 10 mM Tris-Cl, pH 7.4, 1 mM EDTA, 250 mM sucrose, 10 mM NaF, 1 mM phenylmethylsufonyl fluoride. The homogenate were brought to 4°C and centrifuged for 20 minutes at 16,000 x *g*. The 16,000 x *g* supernatant was centrifuged at 48,000 x *g* for 20 minutes to obtain a pellet of high density microsomes and the resulting supernatant was centrifuged for 90 minutes at 200,000 x *g* to obtain a pellet of low density microsomes. The low density microsomes were resuspended at a final concentration of approximately 1-3 mg/ml. Protein was quantified using the bicinchoninic acid protein determination kit (Pierce) with bovine serum albumin as standard.

GLUT4-enriched fractions were then isolated from LD microsomal fractions utilizing the sedimentation sucrose velocity gradient centrifugation (Kandror et al., 1995, Biochem. J. 307:383-390; Heller-Harrision,et al., 1996, J. Biol. Chem. 271:10200-10204). Briefly, 1.5 to 2 mg of LD microsomal fractions were loaded onto a 10-35% sucrose velocity gradient (sucrose in buffer B: 20 mM HEPES, pH 7.4, 100 mM NaCl, 1 mM EDTA, 2 mM dithiothreitol, 1 mM, 10 mM NaF, 1 mM NaPPi, 0.1 mM Na_3VO_4, 1 mM phenylmethylsulfonyl fluoride, 10 μg/ml aprotinin and 10 μg/ml leupeptin) and

centrifuged for 3.5 hours at 110,000 x *g* rpm in an SW28 rotor (Beckman) and 1 ml fractions were collected. The crude membrane fraction contains most of the GLUT4 present in unstimulated adipocytes and is composed primarily of intracellular membranes (Czech and Buxton, *supra*). This additional centrifugation step separates about 90% of the total membrane protein (fractions 1-7) from the GLUT4-enriched membranes (fractions 8-18).

Insulin treatment of rat adipocytes prior to disruption of the cells and preparation of these membranes causes a marked decrease in the yield of GLUT4 present in the latter fractions. However, no such insulin effect is observed when total membrane protein is measured because these membranes are still highly contaminated with membranes that do not contain GLUT4 and are not insulin-responsive.

To further resolve the membrane species associated with GLUT4, fractions 8-18 which contained most of the GLUT4 from the sucrose velocity gradient were subjected to equilibrium gradient centrifugation. Fractions from sucrose velocity gradients containing GLUT4-membranes (Fractions 8 to 18) were pooled, pelleted by ultracentrifugation at 48,000 rpm for 1.5 hours, resuspended in buffer B and then loaded onto an equilibrium density sucrose gradient (10-65% (w/v) in buffer B and centrifuged at 150,000 x *g* rpm for 18 hours in a SW 50.1 rotor (Beckman). After centrifugation, 0.25 ml fractions were collected starting from the top of the gradient. Fractions were analyzed for the total protein content using a Bradford assay (Bio-Rad).

Most of the membrane protein was distributed over fractions 5-20 after this procedure, whereas most of the GLUT4 was distributed within fractions 7-14. Importantly, GLUT4 was localized into two types of membranes (GLUT4 membranes) that can be distinguished based on their sensitivity to insulin. The amount of GLUT4 in fractions 7-9 (peak 1) was decreased when the cells were treated with insulin before homogenization and preparation of membranes, whereas the GLUT4 in fractions 10-20 (peak 2) was not affected by insulin treatment of the adipocytes. Strikingly, measurement of total membrane protein in the fractions of this gradient revealed a similar profile: about a 50% reduction in fractions 7-9 due to insulin action, with no insulin effect observed in fractions 10-20. This observed insulin-mediated decrease in total membranes recovered in fractions 7-9 indicates the successful partial purification of membranes of the insulin-responsive compartment or compartments in primary adipocytes. Similar data were obtained using 3T3-L1 adipocytes.

These methods can be used to, e.g., provide an enriched preparation of glucose transport-related sequences. In addition, in screening assays, a test compound can be incubated with the cells before isolation of the vesicles and the ability of the test compound to affect the localization of the glucose transport-related sequence determined.

Example 5: Characterization of GLUT4 Membranes

Two additional approaches were used to characterize the membranes resolved by equilibrium gradient centrifugation. First, each fraction from the gradient was analyzed by SDS-PAGE and silver staining of the constituent proteins. This analysis revealed that most of the membrane proteins in fractions 7 and 8 were dramatically reduced when membranes were derived from insulin-treated adipocytes. Certain proteins in fractions 6 and 9 showed the same effect, whereas many did not. These results suggest that membranes resolved in fractions 7 and 8 are highly purified insulin-responsive membranes, while those in fractions 6 and 9 are only partially purified. Membranes in higher density fractions show no detectable insulin-sensitivity in spite of the presence of significant GLUT4 protein. Many of the protein bands in the insulin-sensitive membranes are also present in the membranes that are not responsive to the hormone. These data are consistent with the hypothesis that the insulin sensitive membranes containing GLUT4 contain many of the same constituent proteins as other cell membranes that function in a hormone-insensitive mode. Thus, these proteins may also be targets for drugs that potentiate insulin action and ameliorate type II diabetes.

To further characterize the GLUT4 membrane preparation, we determined the distribution of transferrin receptors, thought to be present in endosomal membranes, and VAMP2 (vesicle-associated membrane protein), thought to be associated with insulin-sensitive GLUT4-containing membranes (Kandror and Pilch, 1996, J. Biol. Chem. 271:21703-21708: Kandror and Pilch, 1996, Am. J. Physiol. 271:E1-E14). Surprisingly, both of these proteins were present in the fractions that were responsive to insulin and their distributions were more restricted to these fractions than was GLUT4 itself. These data suggest that the insulin-sensitive membranes in these fractions are contaminated by recycling endosomes, that transferrin receptor is present in the insulin-sensitive membranes, or both. The restriction of VAMP2 to the insulin-sensitive fractions is consistent with data showing that VAMP2 function is necessary for GLUT4 translocation to the plasma membrane in

response to insulin (Cain et al., 1992, J. Biol. Chem. 267:11681-11684; Martin et al., 1996, J. Cell. Biol. 134:625-635).

Expression of transferrin and/or VAMP2 can therefore be used as part of a system analyzing glucose transport, e.g., in diagnosing type II diabetes.

These experiments provide an example of a method for analyzing glucose transport, e.g., in an individual with type II diabetes. In such a case, insulin-sensitive cells from the individual are cultured and analyzed as above. Alterations in the amount or distribution of vesicle proteins compared to a control (i.e., normal with respect to diabetes) indicate that the individual has or is at-risk for a disorder involving glucose transport. Testing cells from the individual that were cultured in the presence or absence of insulin provides additional information regarding hormone sensitivity (e.g., by examining the distribution of vesicle proteins in the presence and absence of hormone.

Example 6: Identification of cytoskeletal proteins in GLUT4-containing membranes

To identify proteins present in the insulin-sensitive membranes containing GLUT4, the equivalent of fractions 7 and 8 were pooled, analyzed by SDS-PAGE and the gels silver stained. These results confirmed that many of the resident proteins in the membranes derived from insulin-treated cells were present at lower abundance compared to controls. Many of the protein bands, combined from both lanes, were subjected to tryptic hydrolysis and the peptides analyzed by mass spectrometry as described in Example 6. Of the proteins identified by this procedure, peptides derived from GLUT4 itself appeared in two closely spaced bands. Remarkably, the lower of these bands also contained a peptide corresponding to the phosphorylated form of the COOH-terminus of GLUT4, indicating significant amounts of phosphorylated GLUT4 are present in insulin-sensitive membranes. In addition, peptides corresponding to several proteins previously reported to be present in these membranes were identified, including the IGF-II/mannose-6-phosphate receptor, IRAP (insulin-regulated aminopeptidase), amine oxidase, long chain acyl-CoA synthetase, and SCAMPs (secretory carrier-associated membrane proteins). Two proteins not previously known to be present in insulin-sensitive GLUT4-containing membranes were also identified—vimentin, an intermediate filament subunit, and α-tubulin, the microtubule protein.

Two approaches were taken to determine if vimentin and α-tubulin are directly associated with membrane vesicles that also contain GLUT4 and are insulin-sensitive. In one approach, the membrane preparations obtained from the equilibrium gradient centrifugation were analyzed by MALDI-TOF MS analysis. In a second approach, the fractions were analyzed using immunoelectron microscopy using anti-GLUT4, anti-vimentin and anti-tubulin antibodies.

MALDI-TOF MS Analysis

Proteins resolved by SDS-PAGE were visualized by silver staining (Bio-Rad) and the bands were excised from one single dimensional 5-15% gel. The silver stained proteins bands were destained and tryptically digested (trypsin) in gel according to Gharahdaghi et al. (1999, Electrophoresis 20:601-605) with some slight modifications. The digested samples were further concentrated and desalted with Millipore Zip Tip C18 micro tips prior to MALDI-TOF (matrix-assisted laser desorption ionization time-of-flight) analysis. MALDI-TOF analyses were performed on a Kratos Analytical Kompact SEQ Instrument, equipped with a curved field reflectron. Peptide masses were searched against the non-redundant protein database using MS-Fit of the Protein Prospector program developed by Clauser et al (1999, Anal. Chem. 71:2871-2882) at University of California, San Francisco. Fragmentation information obtained from individual peptides via Post-Source-Decay (PDS) analysis was searched against the non-redundant protein database using the protein prospector program MS-Tag.

Immunoelectron Microscopy

Standard techniques were used to stain the prepared vesicles with anti-Glut4, anti-vimentin, and anti-tubulin antibodies conjugated to colloidal gold particles. Most of the vesicles in the preparations show reactivity with anti-GLUT4 indicating relatively low contamination with membranes that do not contain the transporter. Anti-vimentin and anti-tubulin antibodies were used to detect vimentin and tubulin in GLUT4-positive membranes. A fraction of these GLUT4-positive membrane vesicles also directly react with anti-vimentin and anti-tubulin. Non-immune antibodies showed no detectable staining of these membranes under the conditions of these experiments, while anti-GLUT4 staining was readily detected.

These results indicate that some GLUT4-containing membrane vesicles are associated with the cytoskeletal proteins vimentin, α-tubulin, or both.

To further assess association of vimentin and α-tubulin with insulin-sensitive membranes, the abundance of these cytoskeletal proteins was estimated using Western analysis in each of the membrane fractions obtained by equilibrium gradient centrifugation. The relative abundance of GLUT4 protein versus vimentin and α-tubulin throughout these fractions was analyzed. Both vimentin and alpha-tubulin are present in all of the membrane fractions of the gradient except for the top few fractions. Strikingly, both of these proteins are greatly reduced in abundance in the same gradient fractions in which GLUT4 is reduced in response to the action of insulin. In membrane fractions of higher density, the concentrations of GLUT4, vimentin, and α-tubulin are all unaffected by prior treatment of cells with insulin. Taken together, these experiments demonstrate that two cytoskeletal proteins, vimentin and α-tubulin, are bound to subpopulations of the GLUT4-containing membranes that are insulin-responsive in rat adipocytes.

Example 7: Identification of Proteins Expressed in GLUT4-Containing Vesicles

GLUT4-containing membranes were isolated by velocity sedimentation, then further fractionated using sucrose density equilibrium gradients, and, as described above, GLUT4-containing fractions that exhibited the most insulin sensitivity (peak 1; fraction 7-8 and the fractions containing GLUT4 that were less insulin sensitive (when compared to the peak fractions) were identified. The biogenesis of the peak 1 vesicle fraction was also observed to increase during 3T3-L1 adipocyte differentiation. To identify proteins present in GLUT4-containing vesicles, fractions corresponding to peak 1 from primary adipocytes, peak 1 from 3T3-L1 adipocytes, and peak 2 from 3T3-L1 adipocytes were pooled, subjected to SDS-PAGE and silver stained. The protein bands were subjected to tryptic hydrolysis and the peptides analyzed by mass spectrometry using standard techniques. Figs. 8A-8I are a list of the peptides identified in peaks 1 and 2, as well as their GenBank Accession numbers and the Genbank Accession numbers of a human homolog if one is available.

These proteins are useful as targets for compounds that modulate glucose transport as well as for diagnosis of individuals having or at risk for disorders related to glucose transport.

Example 8: Comparison of Muscle-Adipocyte Union Library Sequences and GLUT4-Enriched Vesicle Sequences

A comparison was made between the glucose transport-related proteins identified in the subtractive and the Adipocyte Union libraries and glucose transport-related proteins identified in glucose transport vesicles. Fig. 9 lists those proteins that were in common between at least one of the libraries and were also identified in peak 1 or 2 of the vesicle preparation. Acetyl-CoA carboxylase, carboxylesterase, caveolin-1, CDC36, are listed in this figure although their presence in peak 1 or peak 2 is not confirmed.

Example 9: Analysis of Gene Expression Using DNA Arrays

DNA arrays can be used to assay the levels of gene expression of selected gene sequences. These were measured by assaying the amount of mRNA for the gene sequences selected for analysis in undifferentiated 3T3 L1 fibroblasts and differentiated 3T3 L1 adipocytes. The sequences selected for analysis are selected from the MAU library. Clones from the library that show significantly different levels of expression in differentiated adipocytes are selected for further analysis of their role in glucose transport.

A protocol for analyzing an array follows.

1. Clones that are previously sequenced are selected from the MAU library. These clones consist of known and unknown genes with various levels of expression in fibroblasts and adipocytes.

2. Each of the clones is diluted 1:50 and then amplified by PCR.

3. PCR fragments are gel purified and re-suspended in 20-30 μl of ddH_2O.

4. Nucleic acid concentration of the PCT products is measured by spectrophotometer (OD_{260}) and further dilutions are made bringing all samples to a concentration of 100 ng/μl.

5. The PCR samples are then dot blotted (i.e., each to a separate address) onto a charged nylon membrane at 50 ng per dot as described in steps a – c.

a. The PCR samples are diluted to the desired concentration in 0.2 M NaOH/10 mM EDTA (denaturation solution) and then incubated at 37°C for fifteen minutes.

b. The nylon membranes are pre-wetted and placed into a dot blot apparatus. Suction is applied to the apparatus and buffer is washed through the openings.

c. After denaturation the DNA solution is place in the apparatus (each sample in a separate well) and suction is applied. Once the solution has gone through the filter, the wells are washed with additional denaturation solution. The membrane is then removed from the apparatus and cross-linked with UV-radiation. Membranes are then baked to dryness and stored in sealed bags until ready for use. The membrane with the PCR sample is referred to as an array.

6. To analyze expression, the arrays are pre-hybridized for at least five minutes in modified Church's buffer (7% SDS, 1mMEDTA, 0.5 M NaHP04 pH 7.2).

7. Probes for the arrays are labeled in a modified first strand cDNA synthesis reaction as follows:

a. Two labeling reactions are carried out side by side. One using adipocyte mRNA as the substrate and using fibroblast mRNA as the substrate.

b. For each labeling reaction, 2 μg of mRNA is combined with 2 μl of oligo d(T) and 2 μl of random hexamer and incubated at 70°C for 10 minutes and then chilled on ice.

c. After the incubation, add 4 μl of 5X first strand buffer, 2μl of 0.1 M dithiothreitol (DTT), and 1 μl of a modified cNTP solution (A, T, and G at 500 μM final; C at 5 μM final), and 5 μl of labeled dCTP. Mix, microfuge, and place at 37°C for 2 minutes.

d. Add reverse transcriptase (2μl Superscript II; Life Technologies Inc.; Rockville, MD), mix and place at 32°C for one hour.

e. Place on ice to stop reaction.

8. Unincorporated dNTPs are removed from the probe mixture by passing the mixture through a G50-150 Sephadex column (Sigma) and centrifuging for 1 minute at 1000 x *g*.

a. To the labeling reaction add 1 μl 1% SDS, 1 μl 0.5M EDTA, and 3μl 3M NaOH and incubate at 68°C for three minutes and then at room temperature for fifteen minutes.

b. Add 10 μl of 1 M Tris-HCl pH 7.5 and 3 μl of 2N HCl.

c. Add an additional 50 μl of TEN (10mM Tris-Cl, 1mM EDTA, 100 mM NaCl, pH 8.0) buffer to the tube and filter the labeled mix through a G50-G150 Sephadex column to remove unincorporated nucleotides.

d. Add 50 µg of CotI DNA (Life Technologies Inc.; Rockville, MD) to this mixture; boil for five minutes, and hold at 68°C until ready to use.

9. The probe is added to a sufficient volume of modified Church's buffer and the mixture is added to the filters (add approximately the same number of counts to each array) and hybridized overnight at 65°C with gentle rocking.

10 After hybridization the filters are washed as follows: twice at room temperature with 2XSSC/0.05%SDS for five minutes, once at room temperature with 0.1XSSC/0.1%SDS for ten minutes and finally once or twice at 65°C with 0.1XSSC/0.1%SDS for 1 hour.

11 The damp arrays are wrapped in plastic wrap and put on a phosphor-imaging screen overnight (Filters may also be placed on auto-rad film).

12 Commercially available programs for phosphor-imagers quantify images. Alternatively the images can be quantified with commercially available graphics or image analysis programs. The quantified values represent the relative amount of expression of each sequence on the array.

13 The values are further analyzed by subtracting background from each measurement and the values are then graphically represented to facilitate comparisons between the values for fibroblast and for adipocytes.

This method allows for screening of multiple sequences in a single procedure. Such methods are useful for analyzing expression profiles in individuals having or at risk for a disorder related to glucose transport, for analyzing the ability of a test agent or a candidate agent to alter expression of a gene involved in glucose transport, and to analyze compounds that may be useful as drugs for other disorders for potential (deleterious) side effects resulting from unintended alterations in expression of genes involved in glucose transport.

Similar methods of analysis using arrays can be used for diagnostic purposes. For example, expression of sequences encoding proteins involved in glucose transport can be analyzed using a nucleic acid sample from the cells of an individual suspected of having a glucose transport-related disorder (e.g., type II diabetes). In general, the nucleic acid sample will represent sequences expressed in a cell type that conducts glucose transport. The sequences analyzed include sequences more highly expressed in adipocytes and/or muscle cells than in fibroblasts (including sequences expressed in adipocytes and/or muscle cells and having no detectable expression in fibroblasts). Such sequences are described herein. The

level of expression of the sequences represented in the array is compared to a reference level of expression (representing the amount of expression present in an unaffected individual who is not at risk for the disorder). An alteration in the level of expression of one or more of the sequences indicates that the individual has or is at risk for the glucose transport-related disorder. The array may include one or more sequences that are used as standards (i.e., reference sequences) to normalize the data between reactions. In general, the sequences used as standards correspond to genes whose expression is not affected in glucose transport disorders. Sequences used as standards can also correspond to genes that are not differentially expressed between adipocytes, muscle cells, and fibroblasts. Examples of such sequences are described herein.

Example 10: Genechip Identification of Genes Not Expressed in 3T3-L1 Fibroblast, but Present in 3T3-L1 Adipocytes and Muscle

To further identify genes that are preferentially expressed in cells conducting glucose transport, the mouse U74A Genechip (Affymetrix) was probed with two independently produced sets of probes from 3T3-L1 fibroblast, 9 day old 3T3-L1 adipocytes, and mouse muscle. The experiments were carried out using standard methods, essentially as described above. The genes listed in Figs. 13A-13C are those whose expression was not detected in fibroblasts, and was detected in adipocyte or muscle on one or both of the duplicate Genechips based on the Absolute call of gene expression made by the Affymetrix Microarray Suite Software. The columns in Figs. 13A-13C marked f1 and f2 are data from the fibroblast replicate chips. The columns marked a1 and a2 are data from the adipocyte replicate chips, and the columns marked m1 and m2 are data from the muscle replicate chips. A indicates that the gene is absent in a tissue. P indicates that the gene is present in a tissue. An M indicates marginal signal and the software cannot determine if the gene is absent or present. The function classes of proteins listed in the last column are: Class 1 are genes encoding metabolic proteins; Class 2 are genes encoding signaling proteins; Class 3 are genes encoding cytoskeletal or trafficking proteins; and Class 4 are other proteins whose function is something other than those of Classes 1-3; and Class 5 are proteins of unknown function. Genes in italics encode mitochondrial proteins.

Genes that are expressed in adipocyte and/or muscle and are not expressed in fibroblasts are useful, e.g., for identifying genes whose expression is altered in disorders involving glucose transport, for detecting aberrations in glucose transport, and as targets for drugs designed to alter glucose transport. Genes that are expressed in both fibroblasts and adipocytes and/or muscle cells are also useful as reference sequences, e.g., to normalize data obtained when measuring expression patterns of genes expressed in glucose transport in a sample.

Example 11: Probe sets on Affymetrix GeneChip U74A whose expression is increased in both 3T3-L1 adipocytes and muscle compared to fibroblasts.

To determine the relative expression levels of genes in cells that conduct glucose transport compared to cells that do not conduct glucose transport, the mouse U74A GeneChip was probed with three independently produced cDNA probes from 3T3-L1 fibroblasts, 9 day old 3T3-L1 adipocytes, and mouse muscle. The experiments were conducted using standard methods, essentially as described above. The genes listed in Figs. 14A-14G are those whose expression was determined to be the same on all fibroblast chips, and increased on both adipocyte or muscle GeneChips based on the difference change of gene expression made by the Affymetrix Microarray Suite Software when compared to the first fibroblast chip. The columns marked f1, f2, and f3 are fibroblast replicate chips. The columns marked a1, a2, and a3 are adipocyte replicate chips, and the columns marked m1, m2, and m3 are the muscle replicate chips. NC indicates no change of expression. MI indicates that there was a moderate increase in expression. An I indicates an increase in expression. The function classes of the genes listed in the last column are as follows: Class 1 genes encode metabolic proteins; Class 2 genes encode signaling proteins; Class 3 genes encode cytoskeletal or trafficking proteins; Class 4 genes encode proteins with functions other than those of Classes 1-3; and Class 5 are proteins of unknown function. Genes listed in italics encode mitochondrial proteins.

Genes with increased expression in adipocyte and/or muscle compared to fibroblasts are candidate genes for a glucose transport pathway. Such genes are useful, e.g., for identifying genes whose expression is altered in disorders involving glucose transport, detecting aberrations in glucose transport (e.g., for diagnostic purposes), and as targets for

drugs designed to alter glucose transport. Genes whose expression is the same in fibroblasts and adipocytes and/or muscle cells are also useful as reference sequences, e.g., to normalize data obtained when measuring expression patterns of genes expressed in glucose transport in a sample.

In selecting nucleic acid sequences for the uses described herein, any of the genes or sequences identified using any of the above methods (i.e., subtraction libraries, vesicle proteins, or microarrays) can be combined. Particularly useful are those sequences corresponding to genes found to be preferentially expressed in adipocytes or muscle cells compared to fibroblasts in at least two of the methods. In some embodiments, the sequences are selected from those that are preferentially expressed in both adipocytes and muscle cells compared to their expression in fibroblasts in at least two of the methods.

Example 12: Assay for GLUT4 transport/insulin mediated transport

Methods are available for the rapid testing of the functions of proteins identified as glucose transport-related proteins, e.g., by assaying their role in GLUT4 regulation. For example, a reporter molecule that is a chimera of the transferrin receptor (exofacial domain) and the IRAP (insulin-regulated aminopeptidase) protein that traffics in cells like GLUT4 has been described as a surrogate for GLUT4 (Johnson et al., 2001, Mol. Biol. Cell 12:367-381; Lampson et al., 2000, J. Cell Sci. 113:4065-4076; Subtil et al., 2000, J. Biol. Chem. 275:4787-95; Johnson et al., 1998, J. Biol. Chem. 273:17968-17977). This chimera is expressed in cells and is sequestered in the perinuclear region under basal conditions. Insulin then stimulates the chimera's translocation to the cell surface. The translocation can be readily measured using an antibody raised against the exofacial domain of the transferrin receptor or by labeled transferrin itself. This assay is then applied to cells in which the protein of interest (e.g., a glucose transport-related protein) has altered expression. For example, the protein of interest can be overexpressed in a cell that also expresses the transferring/IRAP chimera, and the effect of overexpression on insulin regulation of translocation assayed. This assay can also be used to determine if a test agent or candate agent targeted to a glucose transport-related protein is an effective modulator of insulin regulation of translocation. For example, the candidate agent can be a ribozyme or antisense sequence that is targeted to a nucleic acid sequence encoding a glucose transport-related

protein, e.g., RabGAP or endophilin 1b. Similarly, the assay can be performed in the presence and absence of a candidate agent targeted to a glucose transport-related protein or nucleic acid sequence. An alteration in transport of the chimera in the presence of the candidate agent indicates that it is a candidate agent, useful for treating a disorder associated with aberrant glucose transport (e.g., type II diabetes).

Two examples of genes identified using the methods described herein that can be used in the assay methods described above are those encoding an apparent RabGAP and endophilin 1b. The RabGAP protein is predicted to be a negative regulator of Rab GTPases, which are known to promote membrane recycling of GLUT4 as it transits from intracellular storage sites to the plasma membrane and back into the cell. One such protein, Rab 4, is implicated in directing GLUT4 to its perinuclear recycling compartment, a necessary step for GLUT4 to respond to insulin. The RabGAP that was identified is predicted to inhibit Rab 4 by increasing the GTPase activity of Rab 4 leading to its binding GDP and deactivation. Thus, RabGAP is an excellent drug target in that its inhibition might lead to promoting Rab4, a required element in the regulation of GLUT4 by insulin. Endophilin 1b is related to a class of brain endophilin proteins that are involved in promoting endocytosis of plasma membrane proteins. The high expression of endophilin 1b in adipocytes indicates that it is likely to be involved in endocytosis of GLUT4 in these cells. Endophilin 1b is therefore another potential drug target in that its inhibition by a drug is predicted to retain GLUT4 at the cell surface membrane where it can promote glucose transport, thereby lowering blood glucose.

OTHER EMBODIMENTS

It is to be understood that while the invention has been described in conjunction with the detailed description thereof, the foregoing description is intended to illustrate and not limit the scope of the invention, which is defined by the scope of the appended claims. Other aspects, advantages, and modifications are within the scope of the following claims.

WHAT IS CLAIMED IS:

1. A method of identifying a gene whose expression is altered in a glucose transport-related disease or disorder, the method comprising:

providing a nucleic acid array comprising 4 or more nucleic acids immobilized on a solid support, each nucleic acid comprising a sequence of 10 or more consecutive nucleotides within any one of the sequences listed in Figs. 1, 2A-2R, 3A-3E, 6A-6E, 7A-7U, 8A-8I, 9, 13A-13C, and 14A-14G or a complement thereof;

providing a reference nucleic acid sample prepared from a tissue of a normal, control mammal;

contacting the array with the reference sample;

detecting hybridization of the reference sample with nucleic acids in the array, to obtain a reference pattern of glucose transport-related gene expression;

providing a test nucleic acid prepared from a tissue of a mammal having a glucose transport-related disease or disorder;

contacting the array with the test sample;

detecting hybridization of the test nucleic acid with nucleic acids in the array, to obtain a test pattern of glucose transport-related gene expression; and

comparing the reference pattern with the test pattern to detect a gene whose expression is altered in the test pattern relative to its expression in the reference pattern.

2. The method of claim 1, wherein the array comprises 10 or more nucleic acids

3. The method of claim 1, wherein the array comprises 100 or more nucleic acids.

4. The method of claim 1, wherein the array comprises not more than 100 nucleic acids.

5. The method of claim 1, wherein the array comprises not more than 200 nucleic acids.

6. The method of claim 1, wherein the array comprises not more than 300 nucleic acids.

7. The method of claim 1, wherein the sequence comprises 30 or more nucleotides.

8. The method of claim 1, wherein the reference nucleic acid and the test nucleic acid are cDNAs.

9. The method of claim 8, wherein the cDNAs comprise a fluorescent label.

10. A nucleic acid array comprising 4 or more nucleic acids immobilized on a solid support, each nucleic acid comprising a sequence of 10 or more consecutive nucleotides within any one of sequences listed in Figs. 1, 2A-2R, 3A-3E, 6A-6E, 7A-7U, 8A-8I, 9, 13A-13C, and 14A-14G.

11. The array of claim 10, wherein the array comprises 100 or more nucleic acids.

12. The array of claim 10, wherein the array comprises not more than 100 nucleic acids.

13. The array of claim 10, wherein the array comprises not more than 200 nucleic acids.

14. The array of claim 10, wherein the array comprises not more than 300 nucleic acids.

15. An isolated nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:1-3, or a complement thereof.

16. A nucleic acid molecule of claim 15, consisting of a nucleotide sequence selected from the group consisting of SEQ ID NOS:1-3, or a complement thereof and a non-nucleic acid modifying group bound to either a 3' or 5' end of the nucleotide sequence or both.

17. A nucleic acid molecule of claim 15, consisting of a nucleotide sequence selected from the group consisting of SEQ ID NOS:1-3, or a complement thereof, and a synthetic nucleic acid sequence bound to a 3' or 5' end of the nucleic acid sequence or both.

18. An isolated polypeptide comprising an amino acid sequence encoded by a nucleic acid sequence selected from the group consisting of SEQ ID NOS:1-3.

19. An isolated nucleic acid molecule comprising a nucleic acid sequence selected from the group consisting of SEQ ID NOS:4-93, or a complement thereof.

20. A nucleic acid molecule of claim 19, consisting of a nucleotide sequence selected from the group consisting of SEQ ID NOS:4-93, or a complement thereof and a non-nucleic acid modifying group bound to either a 3' or 5' end of the nucleotide sequence or both.

21. A nucleic acid molecule of claim 19, consisting of a nucleotide sequence selected from the group consisting of SEQ ID NOS:4-93, or a complement thereof, and a synthetic nucleic acid sequence bound to a 3' or 5' end of the nucleic acid sequence or both.

22. An isolated nucleic acid molecule of claim 19, consisting of a nucleic acid sequence selected from the group consisting of SEQ ID NOS:4-93, or a complement thereof, .

23. An isolated polypeptide comprising an amino acid sequence encoded by a nucleic acid sequence selected from the group consisting of SEQ ID NOS:4-93.

24. A method for identifying a candidate agent, that modulates the expression or activity of a glucose transport-related polypeptide, the method comprising:

a) providing a sample containing a glucose transport-related polypeptide;

b) adding a test agent to the sample;

c) assaying the sample for expression or activity of the glucose transport-related polypeptide; and

f) comparing the effect of the test agent on expression or activity of the glucose transport-related polypeptide relative to a control, wherein a change in glucose transport-related polypeptide expression or activity indicates that the test agent is a candidate agent that can modulate expression or activity of the glucose transport-related polypeptide.

25. The method of claim 24, wherein the test agent is selected from the group consisting of a polynucleotide, a polypeptide, a small non-nucleic acid organic molecule, a small inorganic molecule, and an antibody.

26. The method of claim 24, wherein the test agent is selected from the group consisting of an antisense oligonucleotide and a ribozyme.

27. The method of claim 24, wherein the glucose transport-related polypeptide is assayed using an antibody.

28. The method of claim 24, wherein the glucose transport-related polypeptide is a human glucose transport-related polypeptide.

29. The method of claim 24, wherein the method comprises the step of determining whether glucose transport is modulated in the presence of the test agent.

30. The method of claim 29, wherein glucose transport is decreased in the presence of the test agent.

31. The method of claim 29, wherein glucose transport is increased in the presence of the test agent.

32. The method of claim 24, wherein the assay is a cell based assay.

33. The method of claim 24, wherein the assay is a cell-free assay.

34. The method of claim 24, wherein the glucose transport-related polypeptide is selected from the group of polypeptides encoded by sequences comprising the nucleic acid sequences listed in Figs. 1, 2A-2R, and 3A-3E, and the polypeptides listed in Figs. 6A-6E, 7A-7U, 8A-8I, 9, 13A-13C, and 14A-14G 6-9.

35. A method for identifying a candidate agent that modulates expression of a glucose transport-related polynucleotide, the method comprising:

a) providing a sample in which a glucose transport-related polynucleotide is expressed;

b) adding a test agent to the sample;

c) detecting expression of the glucose transport-related polynucleotide;

d) determining the amount of expression of the glucose transport-related polynucleotide; and

e) comparing the effect of the test agent on the amount of expression of the glucose transport-related polynucleotide in the sample relative to a control, wherein a change in the amount of expression from the glucose transport-related polynucleotide indicates the test agent is a candidate agent that can modulate expression of the glucose transport-related polynucleotide.

36. The method of claim 35, wherein the test agent is selected from the group consisting of a polynucleotide, a polypeptide, a small non-nucleic acid organic molecule, a small inorganic molecule, and an antibody.

37. The method of claim 35, wherein the test agent is selected from the group consisting of an antisense oligonucleotide and a ribozyme.

38. The method of claim 35, wherein the glucose transport-related polynucleotide is a human glucose transport-related polynucleotide.

39. The method of claim 35, wherein the method comprises the step of determining whether glucose transport is modulated in the presence of the test agent.

40. The method of claim 39, wherein glucose transport is decreased in the presence of the test agent.

41. The method of claim 39, wherein glucose transport is increased in the presence of the test agent.

42. The method of claim 35, wherein the glucose transport-related polynucleotide is selected from the group of sequences listed in Figs. 1, 2A-2R, and 3A-3E-3 or a complement thereof, and listed in Figs. 6A-6E, 7A-7U, 8A-8I, 9, 13A-13C, and 14A-14G, or a complement thereof.

43. The method of claim 35, wherein the assay is a cell-based assay.

44. The method of claim 35, wherein the assay is a cell-free assay.

45. A method of diagnosing an individual having or at risk for a glucose transport-related disorder, the method comprising:

(a) providing a nucleic acid array comprising 4 or more nucleic acids immobilized on a solid support, each nucleic acid comprising a sequence of 10 or more nucleotides, the sequence comprising or containing a sequence selected from the group of the sequences listed in Figs. 1, 2A-2R, and 3A-3E, or a complement thereof, and the sequences of the genes listed in Figs. Figs. 6A-6E, 7A-7U, 8A-8I, 9, 13A-13C, and 14A-14G, or a complement thereof;

(b) providing a nucleic acid sample from the individual;

(c) contacting the array with the sample from the individual

(d) detecting hybridization of nucleic acid in the sample from the individual with each nucleic acid in the array, to obtain a pattern of glucose transport-related gene expression;

(e) comparing the pattern of glucose transport-related gene expression in sample from the individual with a reference pattern, wherein a comparison of the pattern of expression in the individual compared to the reference pattern indicates whether the individual has or is at risk for a glucose transport-related disorder.

46. The method of claim 41, wherein the array comprises 10 or more nucleic acids

47. The method of claim 41, wherein the array comprises 100 or more nucleic acids.

48. The method of claim 41, wherein the array comprises not more than 100 nucleic acids.

49. The method of claim 41, wherein the array comprises not more than 200 nucleic acids.

50. The method of claim 41, wherein the array comprises not more than 300 nucleic acids.

51. The method of claim 41, wherein the sequence comprises 30 or more nucleotides.

52. The method of claim 41, wherein the sample from the individual is a cDNA sample.

53. The method of claim 48, wherein the cDNA sample comprises a fluorescent label.

54. The method of claim 48, wherein the disorder is type II diabetes.

55. A nucleic acid array comprising 4 or more nucleic acids immobilized on a solid support, each nucleic acid comprising a sequence of 10 or more nucleotides, the sequence consisting of at least a portion of a sequence selected from the group consisting of the sequences listed in Figs. 1, 2A-2R, and 3A-3E, or a complement thereof, Figs. 6A-6E, 7A-7U, 8A-8I, 9, 13A-13C, and 14A-14G, or a complement thereof.

Figure 1

Novel Sequences from Clones in the Muscle-Adipocyte Union Library

Line number 259

>c0148

CCCCAACCTGCTCCATTGCTTGGGGGAGCGGTCCATGAGCGCTTGTCTCATCCCT
GGCCTCCCGGGAAAGTCTATGCAAAAGCTAAGGTTAACA (SEQ ID NO:1)

Line number 258

>c0827

TCACAGAGGCTCTGAGGCTACCACGAAGATGAACTCTCAGAAATGGGATTGTCA
CCCTCGATGAGTTTCCAGTTCCCTCTCTGTTGTATGATGACACAAGAAGGTGAAG
TGTTGCCTCTCTACAACTGGAAGAGGGAGA (SEQ ID NO:2)

Line number 260

>c1083

GTACTAGCGCTTACAGGTCTGTGTGCAGCCATGCCCAGCTTTCTAAGTGGGTGCT
GGAATCTGACTTCAGGTTTCATGCTTGAGCAGCAAAGCCCTCTTACACAGAGCCA
TTTCGACAGTTCTGTGACTTAGGTAGACTCACATCTGTCAGGCTAGAATTTCCAA
AATTGAAAATGAATTCAAAGTGAAATGCTTGGGAAGTAAGTTAAAGATAGGCTA
AATGGTTAACCCAGCAGTTAGGGTTGCTTTCTGCTCTTCCAGGGGACCTAAGTTT
GGTATTTTCTAGCACCCACACTTGGTGGCTCAAGCCCTCTAACTCCAGCTGCAGG
GGATAGGATGCCCTCTTCTAACCTCCACTGGTGAGAATATCTCACACACACACAC
ACACACACACATGCTCACATACACAACCT (SEQ ID NO:3)

Figure 2A

Muscle Adipocyte Union Library

Number: 172
>C0025
GTACACTGACTTGCCCCGTGGGACCATCTCTGAAAACCAGGCTGTGGGGGCT
GGAGACCCTGCCTCGGATCTGAGAGACCAGGCTTCCACATAAAAACGAGGTG
CTAATCTGGCCTTATTAAGTCTGGGCCCAAGTGTTTTCTCTCTCAATAAAATG
ACTTAAGGTAAAAAAAAAACCTC (SEQ ID NO:4)

Line number 229
>C0039
GTACTATGAGCTGAAGATTGCCTTCGTCATTTGGCTGCTGTCGCCCTACACTA
GAGGGGCGAGTTTAATCTATAGAAAGTTCCTTCATCCCCTGCTGTCATCAAAG
GAAAGGGAAATTGATGATTATATTGTCCAAGCCAAAGAAAGAGGCTATGAG
ACAATGGTGAATTTTGGACGGCAAGGTTTGAATTTAGCAGCTGCAGCCGCCG
TCACTGCAGCAGTGAAGAGCCAAGGAGCAATAACGGAGCGTCTGCGAAGTTT
CAGCATGCATGATCTGACAGCTATCCCAGGGGATGAGCCGTGGGGACACAGA
CCTACCAGACTTTG (SEQ ID NO:5)

Line number 240
>C0076
GTACAGTCCATGCTCATCTGAGAAATTTACAGACTACAGTTGGCCAAGTTCCT
CACCATATGGTCATTATACCTTCCATAAGATTTGATTCATGCTTACTTTTCTG
TATCCATTTCTGGCAAACAAGTTACTTGTATCATGACACAGGAGGATTTTAGT
TAGCTCTCTTACACACTATTTTATTGATGCATTATGAGATTTAATGACTATGA
AGGGGAATGATAATTCTAGTTGGCCATCATTGGCAGCACTTACTACTAAAGT
GGAAGTGAGACATTTGGACTGTATATCTGTTTGGTATGTTATATAACTACTAA
AATTGAATGGTGGCAGAAAGGAATGAA (SEQ ID NO:6)

Line number 191
>c0089
GTACCCACGGGATGAAGAACCTTTCATTTCCCCCCCTTTTCCTTTTCTTCCTTG
TCCCTGAACTGCTGGGCTCCAGCATGCCACTGTGTCCCATCGATGAAGCCATC
GACAAGAAGATCAAACAAGACTTCAACTCCCTGTTTCCAAATGCAATAAAGA
ACATTGGCTTAAATTGCTGGACAGTCTCCTCCAGAGGGAAGTTGGCCTCCTGC
CCAGAAGGCACAGCAGTCTTGAGCTGCTCCTGTGGCTCTGCCTGTGGCTCGTG
GGACATTCGTGAAGAAAAAGTGTGTCACTGCCAGTGTGCAAGGATAGACTGG
ACAGCAGCCCGCTGCTGTAAGCTGCAGGTCGCTTCCTGATGTCGGGGAAGTG
AGCGTGGTTTCCAGCACAGCCACCCGTTCCTGTAGCTCCAGAGATGTTTGATG
TCCTCCGGTCTCTACAGGCACCTGCACTCACGTGCGCGAATCCACACACAAG
CACACATACTTAAAAATAAAACAAAACAGGCTCAAAAAAAAAAAAAAAAAA
AAAAAAAAAAA (SEQ ID NO:7)

Figure 2B

Line number 254
>c0095
GTACTTCTGCATGCACTCTTGCATGGCCCGGAAACTGGTCTATACAGTCTGAC
CCCTTGATATCCTCTGTTGCTGTAGTGGAAGCAGGAGAATGCATACTTGAACT
GCTCCCCACAGGGGCCGCTGGCCATTCCCCCAAGACATGGACAATTCCAGTT
TAATATCTCCGTTAGGCAGTATCAACCCGTTGCTCCTCATAGGG (SEQ ID
NO:8)

Line number 200
>c0103
GAGGTACAACATTGTCGGCTTGCGCAGCAACGTGGACTTCCTGCTCCGGCTCT
CCGGCCACCCAGAGTTTGAGGCTGGGAACGTGCACACGGACTTCATCCCTCA
GCACCACAAGGACCTGCTGCCGAGTCACAGCACTATAGCCAAAGAGTCTGTG
TGCCAGGCAGCTCTGGGGCTCATCCTCAAAGAGAAAGAAATGACCAGCGCTT
TTAAACTCCACACTCAAGATCAATTCTCTCCGTTTTCATTCAGCAGTGGGAGA
AGACTGAATATCTCTTACACCAGGAACATGACTCTGAGAAGTGGTAAAAGTG
ATATAGTCATAGCTGTGACGTATAACCGCGATGGGTCATATGATATGCAGAT
TGACAACAAGTCC (SEQ ID NO:9)

line number 193
.>c0121
TGACTCCTCTTACTTGTGAGAAACAGGCACAGAGAAGCCCTGATGCGGTGTC
ACACATGGATGGCAAGGGGCTGAACTAGGTCTCCTGAGGGGGAAAAAGAAA
CATCAGGCAAGGCAACCGTCTGCCTTCACCACATGACTCCTCTTGGCAACCC
AGTGTCTGGGTTGTTGTAGGGAATTACTTTAAGTTATCCAACAAGCCCTAAGC
AGAGGGTCTGTTTCCTTTAGCCTAGATCCTAGAAGAGGCTGGGCTCCACTTGC
CTCCTAAAGCAAGTTTGTCCTACTCCTTGTCTTCTGCTTGATTTCTGGAACTGG
GCTTTGTTTCAAGCATGTATAAAGTTTTGGCCTTACTCCCACCCAAACCCTTTT
ATTCCACCATTCTACAAATTAAGGGAGG (SEQ ID NO:10)

Line number 195
>c0124
GTACTGAGCTTGCCCTGTGCCTTGTGGGGAAGGGAGTGAGAAGGCAGCATAA
GAGGTCTTGCCAGTTGAGACAGAAGCAGGTAGGTGCAGGAACTATTTTGGAT
AGGGCATGTTGTTCAGTTGATGGCAACTAATTCTCTTAAGAACATATTGTCTC
CTATTAATGTTGGGGACAAACCCAGAAAGGGAAGGGTATAGAGATGGGGAA
TTGGTGAGATCTTCCAGCCCTGAGACCATGTCTGAGCCACCCCAAACTTAGG
CTATCTCCCAGCATAGGAGTGAGACACTGCCCAGTTTACTCAAATTCTAGTGT
AC (SEQ ID NO:11)

Figure 2C

Line number 113
>c0139
GTACCAAAATATGGACAGAAAAGATAATGAAGCAGACCGAAGTGCTGTTGC
AGCCAAATCCAAATGCCAGGATAGAAGAATTTGTTTATGAGAAACTGGATAG
AAAAGCACCAAGTCGTATAAACAACCCGGAACTTTTGGGACAATATATGATT
GATGCAGGAACTGAGTTTGGCCCAGGGACAGCTTATGGAAATGCCCTTATTA
AATGTGGAGAAACACAGAAGCGAATTGGAACAGCTGACCGAGAGCTGATTC
AAACATCAGCCTTAAATTTTCTCACTCCTTTAAGAAACTTTATAGAAGGGGAT
TACAAAACAATCGCAAAAGAAAGGAAACTATTACAGAATAAGAGACTGGAT
TTGGATGCTGCAAAA (SEQ ID NO:12)

Line number 244
>c0152
GTACATGAACCAGATGTATTTCTCAGCTTTACATAGGGGAAAGGGAATTAAA
AAAATACGCAATTGCCCAGCAAATGCAAATGTTTAAAAAGGAAATGCAGAG
AGAACTATGGGAATGGAACAAACAACAGACAGAACTTCAAACAGTGAAAGA
AAAACAAACAAAACAACCAGAGGGAGAAAAACACAAAAGATCTGAAATCCA
CCAATCGCTTTTTGAGCTGAATGGGGGTTGCTTTAAGACCAGAAGTCAAAGT
CACTGCTGCTGGTGGTCTGCCACGTGGGGGTAGTTCACCTAATTCCAAATAGC
TGGCCCTGCTTCAGGGCTGGGGCACCC (SEQ ID NO:13)

Line number 208
>c0196
AGGGGTGTAAACAGTAAACTGCTTTATTGAGACACTGTTACAACGATTCCTTT
GTTACACAGTTTTAAAATATTTTATAACACTCTTCCTGGGGAGAAGTTAAAAT
CTGAGGCTTAGTTTAGACTGCTGGGAAATATACAATGTAC (SEQ ID NO:14)

Line number 196
>c0222
CAAGTTCTTCAGTTACAACCTAGTAGTATACTTACTCTTCCAACTGTCCTAAG
GTCACTTCCCAGCCAGCTTAGGATCTTCAGCATTTTTAAGAGCTGAAGCTCCC
TCTTGCCCTTCTTGTCTACTCCTCACTGCCAGTTGGGGCCTAGGCTTAGTCCTG
GGCAAATGTCCATGATCTTGCTGCTGTAGGAAGCTTGATAGGGCATTTGGCTC
AAATTTCAGAAGGCCTCGCTCCTGACCTAATTTCTCAAAGCTCCGGTAGTTCT
AGAACCCTCCAATTTCTCATCTGGTTGCAAGGCTTATTTTTCTTTT (SEQ ID
NO:15)

Figure 2D

Line number 243
>C0236
GTACATTCAACTTCTGCCCTGTAATTCGGCCAAGTAAGGCCCATATCCCTTGC
CCTTCACTTCGAAGTTTCCCATTCAGATTTTGCAGTTCCTCTAATGATTCACAG
AGCTTATTATATTCTACATGAAGCTTTTCTTGCTCATTCTCTAATTTATCCTTT
ATATTCTTTTGTTCAGCCATATTTTCCTGAATTTGAATCATGCGCATATTTCTC
TCAAAATAGTAATTCACAAAATCAGCAGTGAGAGCGGGCTGCTTATGCTTCC
GCCTTCCATCCACACTAGAACTAGTATCTGAATTTGTCCACTGGAAAGATATC
CGT (SEQ ID NO:16)

Line number 203
>c0250
GTACAAATGTCCAATGCGTATTTTGTGTTTGTCTTAAAAATTCATCAACTCAA
CCTGAATTTAGGAATATACTTTAATAAAGCATTTGTGCGACACAGAAGCGAC
ATGTCTGATAAACTCATAGGAGGATTGTCCCTTGAGGGGAGGGACCATCTTG
AGAATTCACAGTAATAACAGCTATACTGACTGTTAAAACAAAAATTACTGTT
CAATACATGTCTTATCTCCCTACCCTACCCCCGGAAAACAGGAATGGGG
(SEQ ID NO:17)

Line number 167
>c0352
GTACAAGGACCAGCTCTTGAAAGAGACAGTGCTCCAGCCACTGCTGCAGCCA
CAGATCATGTCAGCATGAGTAGTCGTGCTGAAGGGAAAACACAGAATGCTAT
CTTAATGACCATGCCAACATTATTGAATAGCCGAAAGTCCCTAAACCCACTCT
CTGCTGCCTTATCAATGCTAAACCTTATTTGTCTTCATCAAGAGTAGTTCAAA
ATATGCAACTAATTTAATAATTTTGAATGATGGTTTTATCTATAGCAATCTGT
AGTAATATGTATATTATCTATTGGGATTTGTGTAATAAAAAATCTAAGGGAAC
AAAATTTTATAACTACAAGCACTTAAGTAC (SEQ ID NO:18)

Line number 251
>c0367
GTACATATTACCCCGTTTCCTCCTCAATAGAATTGCCACAAACTGCATGCTAA
ATTTAGTTCTTCTGGACAGACCACAACCCTAAGGCTAGTTCTGCTATGTCATA
TATGAGTATTAAATATGGTATGCTTAGTATACTCCAGCCTAAGATAGTTAACC
ACCTGAGACCAGCTGTGATGTTCGAAGACATACAGGATGAGGTTTCTTTCAC
AGGGTTCTGAGCATAGTTTCTGTCCCAGGAATATTGTCTTATCTCCATAACTA
TAGCTGATGCAGAAAGTCCAGACAATATACTCATTTCGACTCAGAATATTTCA
AATTTAGCAATAAACAGTTAGCTTTAGTT (SEQ ID NO:19)

Figure 2E

Line number 197
>c0380
GTACATCTGAATTACACTGTAGGCTCACAAGCAAAAACGTGGTCCCATGGTG
AACAGATGACATCTCAGAGTTCTAAATCTCCACTTTATTTTTAAGATTAGAAG
TTCTCAAAGCTGAGGCTGTAGCTTGGTGGATAGAGAGAGCACTTGCCTAGCA
TTGACAAGACCTCAAGTTCCATCCTCAGCACTTGGGGTGGGGTTATTATTTAT
ATCACTACCATGAAATTATATTGTTTTCATAAAGAAAAATACTCCTAAAATAG
ACTGGGTAAGGCATTGTGGCACACTGTGTTCCTAAGCCACCCAGGAATCTGTT
GAAGG (SEQ ID NO:20)

Line number 236
>C0439
ACCTTGTGCGTTTTTGACCCACAGCCTAGAGAAGGCAGCCACTGCTGGTTACA
GATGGCTTAGTGCAACACAGTGGAGGTAGCAAACATGTAAGGTCCCTAGGGG
CTGTGTGACAGCCTGGAAAGTAAGACACTGGGGACAAACGCAGGGAGGGGC
CAGGCAGCCATTCTTTCTCACCGCCTCCTTTAGAATGTTCCCATAAATTACTA
TTAAAAAAAAAAAAAGTATCCATAGAATACTGAGAGGA (SEQ ID NO:21)

Line number 199
>c0442
GTACTCACCTAATCACTTGTTATCCAGTGCCTGTTCTAGGTTTATGGACTTAA
CTATTTCTGTGATGTTTCATTTTTAGCCATGTTAACTCCTAACACATATTCTCT
TATGTCTCAGTAAAGTTTCATTTGATAAGTTGTTGAGATTCTGTTATTTGATAA
TATTCTTCGGCTGTCCATCCAGCATCTTAATCACTTTAAAACTGTGATTGTTAT
TTGCAACTCTGTTCTTTGGAAAGAATAAAAGCATTTTTTTTCACTTGCTAACA
TGCTCACAAATGTGAGAGAAGAGTCATTAAAAGCTTTACTTACTGGGAAAAA
AAAAAAAA (SEQ ID NO:22)

Line number 198
>c0443
GTCCTTGCCCATTACCAAGAAGTATATGCCAGAGAATAAGGGCGTTCCTCTG
CAAGGGAGCCAGGAGGACAAACCCTTCCCAGACTTTGACCCCTGGTCATCAT
ATAATTGTGAGCAGAACGAGTCATAGCCCATCCCTGCCAACTGCACTGGCTG
TGCCCAGATATTACCCCTCAAGGTAACGCTGCCAGA (SEQ ID NO:23)

Line number 210
>C0504
GTACATGTGCACTAGGGACACACTCCATAAAGGCAGCCATGCTGGGAAAACA
CTTAGCTCCATAATGGTGCCACTCAACTCCAGCTCGCTCTTTTTCGTTCTGATT
GACCGTGATGAGTTGAAGGGTGAGAAACACGGCACCTAGGTGCTGCATGTTT
CTGACATTCTCATAGTCTGTAGAAAATAAAAGTTTATCACACCATCTAAAAA
AAAAAAAAAAA (SEQ ID NO:24)

Figure 2F

Line number 250
>C0507
GTACAGAGAAGCGTGTGCGACCAACCTGCAGTTACAGCTTTCTTCACATTTCC
CCATTCCACACAGCTGTTCCAATGAGCGATCTTGCCCTGGTCACACCACTTTT
AAGCACACTCACAAAACAACCACATATACAGGGTTTTCACAAATATGCAAAT
AAACAGGGTTTTCAGAACCAATTTTTACAAAGTGTTATCTAGTAAGATGACCC
AGGGCCCAGTCTCTTTTATTTCCCAAACTGCTTTGACTGACAGTTTAACAGCT
TCCCCTCTGCTCAGGGATGGTTTTTAAGCACCATTCAAGCCAAACCCACAATG
GATGTATTCAGTGGCAAAGCACTCATGAGGTTTCTACAAGCATCACAATGAT
CTCATAGCAAGCGCAAGACCCTGTCCCCTTGTCACAATGGCTGATTTGAAAG
GTTAAATAAAGCAAACCCTGTTGTCTTCTGGCCAATTCCAGAAATTCAAGATG
TGTTTTGAAGGCATGGCTCACACATTAAATTCCTGTATGTGTTGGATAAGGCA
TGACTTCCATATGAAATTCCTTTTAGATAGCAGCATTTGGAATTCTT (SEQ ID NO:25)

Line number 233
>C0513
GGCTTGTGAGGGTGACTGGACTGTCCAGTGTATGTCTGCTGCTTGTTCCACTG
ACTCTACAGAAACAAAACCCGGTGGGTGCTAGACAGTGCTGGCTCTGCTTGC
AAGCCATGTTGCTGTGGTTGTTACTTATGTGTCTGGTCCAAATAAAGGCAGCT
GCTGATTTGTTGTTAAAAAAAAA (SEQ ID NO:26)

Line number 234
>C0533
GTACAGTTTCATTTAGAAACTTTGTGACCTTGCCCTATTGTTGGGTTTTCTTAT
TTGTCCCTCCAGCTACTTTGTAACAGCCAGAGGTGACCCTTGAAGGAATCCTG
AGAAAGGAGCAAACAGGAATGGAGCCTAGCCACGGTCCCTCAGTTCGGCCTC
AGGGGCGTAGTCCTTCATTGGCTGCATTTTTCTTTGTGCTGGATCACACCCTTC
TGGATCAGATCGGGGACTTCCACTGCCAGCCAAGGACCCAGCCCCAGGGCCA
TGAGATGAGCTAGTCCAAACTTAGGCACATTCCTGGCCTACAAAGGTTTGAA
ATGATCAGTCAGACATATTTTGCCACCCCTGTAC (SEQ ID NO:27)

Line number 231
>C0536
GTACTGCTTCCTGCCAAGAGTGAACGGATCTGTTCTCACTGTGACCCTCAATT
CTCTTCCAGAATGTTCAGAACAAGAGTAAAGTGCGTCACACAGAAAGGAAGT
GCAGTGTGAGTGTTCATGAGTTTGGATCAACTGAGGTCTTTCTTACAAAAAAA
TCATGAAATGTAGCAACCTTTACATTATATTCAATTGCCAATCATGACAGTAT
TTCAGCTCTACCTTCATGATAAACTTTTGTTACTAAAATTTATATACACACAT
AAGTAAAGTTTTAAGTATTTTATGCTGTTTAATGTTGTGTGGCTTTTATACAAT
CTACAGAATAAAGAAAACTAAC (SEQ ID NO:28)

Figure 2G

Line number 230
>C0538
GTACAATATAACAACATCCATATTAGAATGAAAATCGGAGTATTTAAAAGAC
AACATTAAATCTCTCTTTTTTTTTTTACAGCTTGTATTTACAAAATGAATCAAA
ATATTTTTTTTCTTTAAATTCAGGACTCAATAAAATAAACAGCTGGGAATATG
CAATCTACGACTGTGTAGAGAGGACTGAAAAGAAAACCCAATGTCAGCGAA
AAGTGGGGTGAGACTCTAGTAAAGATCATTAGCAAAGTTTTAGGTCCATACT
ATGGTTTTTAGGAATCCCATAAATGGTTTGCATTTCTAGTGTGACAGAAAAAG
AAGCTTTACATCCTGT (SEQ ID NO:29)

Line number 238
>c0564
GTACAAAGGATCCAGAAAAGGGGACATTTTAAAAAAATCACTGTTTAGGCCA
CCAAACCAGGGGAATCGTGGTGTTACCATCCCATTAGAGTCCCAGTTTACAG
GAAGACACACGAGACAGAGAGCAAAGCCAGTCTCTGTGCTTAGCTACTGTGG
GGGACTGAACACATGTAAAGAAGTCCTACATGTCTGCTGGTCCTTTCCCACCA
GCCCCCATCTTCAGGTCTCCTGAGCTGAACTTCAAT (SEQ ID NO:30)

Line number 253
>c0572
GTACCCAGGTCACTGGCTTCATGTTCAACAACTTCTACCTACAGTAAAAGTTA
AAATTAACACCAGGTATCAGCATATCTTTTATAATACGTGCAAATCACATAGT
TTTCCCCATTCTCCCTGCTTTGATACATACACAAGCTGCTACAAAAATAATGC
GAAAATGTTGAGACCAATCTACTATACTTCAGAAATCTCATCAACCCAATAT
ACAAAGTCAAGATAAAAAAGCAGGCTGTTAAGTTATTCCTGGTTCTGACCAT
ATTATTTCTACAAATCTCATGTTCTCAAATTAGAGCTATAAGTTCCAGCCAAT
GTCTCTGTTGTGGGAAAAGCATAGTCCAGCACTGACGTGAAGGGAGGCAAAC
CTGTCCCTTCAATGTCAAGCTCAACGAGGGTGGCGG (SEQ ID NO:31)

Line number 249
>c0580
GTACCCTGAACCTTGAATCTCAGAAGGAAGTTCAGGCGGTGGGTGCCTCCAC
CCCCATCTAGCCTTCCTCACTGGTTCCCGGGGCTGTCAGCGCGGCTCTCAGTC
TTTTTGCACTAGGGCGGGCAGAGGTGGGATGCGCTCCTTAGATGCCTCTCCAG
TATTCAGTGTCCACACTTCCTGGGATAAATCCCACCTCTGGGTTTTCCTTTTTG
GGGAGAGGGGCAGGGCAGGAATTTGAAGCTGTCACTTCCTTGTGAAGTGTGA
ACACAGTTTAGCCATAGGCAGTTGTGATGTCCAGTGTTTGGGATCTATGGGCA
AGTAGTTTTCCCAAAGCAGGCATTCAGTTCTGGAGCAGGCTTGACTCTGGCCA
AACCCTGAGCCACTGTGAATTTGCTGGTGGTCTTCCGACTTACTGGAGTGATT
TTTTAGGTTTTCCAAATCTTATATGACTCTTTTTTTATCTCTAGCTGATAGAAC
CGCATGTAATTAGGTGTAC (SEQ ID NO:32)

Figure 2H

Line number 255
>c0582
GTACAAGAGTTAACAAATTACAATTAAAACGGAATTTGTTATGGCAATTCCA
CAGAACTTAAAAACATGCAACACTGCATGGTAAAAACAGCTTCATTCATATA
CAAAAAAATATTCCTTTCCGACTCATGACAGTGCCTGGAAACTTCTGACATAA
GCTTTTTGCAAAGAATATTTTTAAAAATGTAACGATCTGATTAACCAATTAGT
GCAGTATTAGGAAAGATAATAAACATTATTAGTAAAGAGGTTACAGTGATTT
ATACCAGGTATAGACAGGGCTCAATGTAGTCTCATTAAATAAATGTTCAGTT
AAGAAAATAGTTTTGAAAAAGATCTTATATTGAAGCCATGTATTAATTTTGTT
GAATCAGCTTATATAAATCAATAGTCAAGTTTATTCAGTTAAAGAAAATAGG
ACTATGCTTTCTTATACTCATAAATAGTAC (SEQ ID NO:33)

Line number 194
>c0591
GTACCGAGCTCGGATCCACTAGTAACGGCCGCCAGTGTGCTGGAATTCGGCT
TCGAGCGGCCGCCCGGGCAGGTGTGAGAGGGAACAACTGCCAGGGAGCTGT
TCCAGGGACCACACAGAAAAAGGCCTCGCTAAAGCAACAAACCTGATCATTT
TCAAGAACCATAGGACTGAGGTGAAGCCATGAAGTTCTTGCTGATCTCCCTA
GCCCTATGGCTGGGCACAGTGGGCACACGTGGGACAGAGCCCGAACTCAGC
GAGACCCAGCGCAGGAGCCTACAGGTGGCTCTGGAGGAGTTCCACAAACAC
CCACCTGTGCAGTTGGCCTTCCAAGAGATCGGTGTGGACAGAGCTGAAGAAG
TGCTCTTCTCAGCTGGCACCTTTGTGAGGGTGGAATTTAAGCTCCAGCAGACC
AACTGCCCCAAGAAGGACTGGAAAAAGCCCGAGTGCACAATCAAACCAAAC
GGGAGAAGGCGGAAATGCCTGGCCTGCATTAAAATG (SEQ ID NO:34)

Line number 205
>C0596
GTACATAGTGCAACTTCATCACAATGTCGAGGTTTAACACACATTCAGTGCGT
GTTTTCTGTGCAAGTTCTTAGTGGTCTGGGACCTAGTTGAAGGTATTCATTTG
CAGATCCACGAGATTCAGGAGGTTTGAAGAAAAGAAAAGTTTCCTTGATTTC
TTAAACCAGCACTGGGTGTGCAAAAGGCAGCGGGCAATCCGGCTGTCAACTG
GAGCCAGATCAGGGCGACGTTAAAAGTTAAGATCCAGCCGCGCATCTCGGCC
TCTGGGCTGAGAGCGAGCTAAGTGAGATCCAGGCAGTTCGTTACTTAGACTG
G (SEQ ID NO:35)

Figure 2I

Line number 209
>C0600
GTACCAAAAGTCTTTATTTTTGGATGTTGCCTGAGATGACCTCTTTCCCCACTC
TCTGATTCTCATTCTTCCATGACACTCAGTTCCTCTTGCATACTCCCAAAGATG
GCAAGGTCTAACCATACATGGATCCTCTGGCTTTGGTTTCTGCGTTAAGTAAT
ACTGGTCACTCAGCCTCCATCTTGGAGGAGGATGTTTTCCTATACCACTCCTA
AAGGTAGTGACCGCACACCAAATCTTAGGTCTCCACTGTCTCAAGGGAATAT
AAAGACCCCAGGGTGGGCTGCAGGTTGCTTTAAACAAGGGGACTACAAGGA
CACCCACAGGTTTCTGCTCTACCCTCTGGCTTGGACAGCAACAGAGAGGCTTC
ATCTAAACTTAAGAAATTCTTTTCTACAAGGCAAGCTGTATCTGCTACCACCA
AC (SEQ ID NO:36)

Line number 245
>c0635
GTACAAGAGATGCACACTTGGAAACCTTGTAGTTGTGAGTAACATATTTATA
CCTATCTTAGTGGGCTTCAGAGAAACAACACGTTATGTATAATAAAAATGGA
GAATTAATAGTTTTTCCAGGTATAAATGTAC (SEQ ID NO:37)

Line number 213
>C0681
GTACAAAGGTAACAACTTAGAATCAAGGCCCGATGTAGGGACTATCATATCT
GCCCTATTTGCCAAAGGGAAACATGATAAGCTCCCACCGGACACGGACAAGT
TACTGAAGGGGAGAGCAGCTAGCTTTCCACGTTAAAAGCATCTAAAAATAGC
CCTTATGTAATGGAGCCAGGGGAGGGGTGCACCAGATTAAATCTGGAAGCTA
GAACTGGTAGGCTGGAGCTGTCCTTCTGGAAAATTTGGAAATTTTATAACTAG
CAATCAGACAACTTTATTCTTAAGCTTTATTTAAGTGTGGTAGGAACAGCATA
CGCTATAGTGTAC (SEQ ID NO:38)

Line number 212
>C0684
GTACATTCCATATTGGACAAAACTTATACTTTGAGTTTCATATGAAATTGAAA
ATTGTATTTTTTTCACTGTTTCATTTTTACACATCTCTATAAGTATTATTTACA
ACCTTGAATAAATAAGCAATAGATAATGGCGAGTAAATCAAGTCACAGTAAA
TACAGCTGTTTTTAAATATTACCTTTAGCTACTATTGTATATAATTTAGAATTC
ATTTTTTCACCGACCAAGTGTTTTACTATTTCCAATCAGTGTTGCCTGCAAAG
ACTTTTGTTTATGTGATGTAC (SEQ ID NO:39)

Figure 2J

Line number 211
>C0689
GTACTTTGGGTTTGGGGACCTTGAGCACTGACGTGAAACTCTAAGGAACCTG
AGCCAAAGAAAACTATTCTCTCAGTGCTGGAACTCAGCGAGCACATGTGTAT
TCTGTGCAAGATGCGCATAAGTCGGGAGAAAGCTTTTACAGAGCAACGTTTC
AATCCTGACCAGGGCTCCAACAGACAAGATGTATGTCACGGCGAACAAGCCA
ATAAGCAGCACCCGTCATGGTTTCTGCTTCGGTTCCTGCCTTGAGTTCCTTTG
AGCGTCTCTCAATGGACTGTGGCCAGGGATATGTAAGCCAATAAACCATTTC
CTCCCCCAAGGGGTTTCAAGTCTTTATCTTAGCAATAGAAACCT (SEQ ID NO:40)

Line number 241
>c0692
GTACAGTGTGATAGGACCACTGTCCTGGCAGAGCAGGACTAAAGGAATGTAG
ACCCCATCTCAGTGTCTTCAGAAGGACTGTGAACAGAGGCAGCGTCTGCCAT
GAAACCAGTTCACCGCTCCTCTCCATGAATTGACCTTTTCAATTTTTTTAAAGT
TATTTTTGAGACAGAGCCTCATGTAGCCCAGGCTGGCCTTGAGCTCTGTATGT
ATCTGAGGCTAGTCCTGAATTCCTAACTCTCCTGAGTGCTGAGATTGTCAAGC
ACTGCATTAAACCTGGCTGGTCTCTTGTCTTGACTTCCTATATCTTTGGCTCCT
TCTGTCTCAGAATGATGGGAAGCCTGACGGGGAC (SEQ ID NO:41)

Line number 257
>c0710
GTACTTCCCATGTCGGTTAAAGTGTTTGTAGAGATAGCTGATCCGTTTGTTAA
GGTTATGCAGGTCCTCAATGAACATCCTGTTTCCCACAAATTTCTTCTTTCGG
ACACAACGCTGCAATTCATTTCCTCTCCAGCCTCGAAGCCATACTGGGCCCTT
GATCAGTTCTTTTGGGTGCTTCTCGAAATTCCCCAGGATTCCAATGTTGTCGT
AGACGCCAAACAGAGCCCTCTTCTCGTTCCAGCGATCAACCACTTTGGGAGG
TGGAAGAGTGAGCCTTACGAAGGCCAACCGTGGGACGCCCAGAGAGAAGCT
TCTGCATGCCCGAGGGCACCCAGCCCCACAGTCTTCTACTTGCCCGCCAAGG
GACACTTCCTGCCATTACTCTAGCCTGTCCGCTCAGTAGGGGCAGCAAACCA
GCCCGGAAAAATGGGGAGCGGGGTGTGAATTTCAAAGCCCAGACACAGATT
ACAGTCCCCAGGCGGGGACTTACACGGTAGTGACAGTAAAAAGGAGCCATTT
T (SEQ ID NO:42)

Line number 232
>C0724
GTACGTGGAGCATAAGCAAGGAACTCGGGGCACTGTGAAGCGGGCAGCTCCT
CGCCAGCTGGCCCCAGTCAGATACACTATGGTGATCAGTTCTCTTTCTGCGTA
GATGACAAGAGAAAGATGAGCCATGCTGACAGAGGATGGGGCGGGCGGGTC
CAGCCACCCGTGAGACAAAAAGGCTGACCTCCATGTTGCAGGATTTGCT
(SEQ ID NO:43)

Figure 2K

Line number 192
>c0729
GTACCCCAATTCCTCTGCATCCTGTGGTGATGGGGGTGACTGTGAAGCCCTGC
TGGTCAGGCTTCCTCCAGTGACAGGGGCGGAGTCTGACGAGACTGGCCTAGC
TCTCTACAGAGGGAGCCCCGGTGCTGTCATGCAGGGACATGACTCTTAACAT
GACTCTTACAGGCTTCACAAGACAGTCTTAACTTCTGATCATCCCACCCCTGG
GTTTTAGTGTTAG (SEQ ID NO:44)

Line number 252
>C0739
GTACTTGCTGCCTTCCTCGCGGAGCCGCGTCCTGATGACCTCGTGTGGGTAAG
CGATGCAGGAGGCACATCCCTTAGAAACAGCAGCAGCTGCCATGAGTCCAAA
GAAGCCAGAGGAACTTTTCTCAGCTCCATCTGTGGAGGAGACGATCGGAGCG
TCTTTCAAACACTTCTTTAAGCTCTCATAAATAGCAAAGCAGATGATTGTCTC
CGAGATCCCAGCGTAGGAGGCGGTCAGCCCTCTATAGAAGCCGCGGACGCCT
TCTGTCTGGTAGACACGTCGAGCACACTGGAGTGTGTTCATCTGCTTGCAGCC
CCTCACCTTGCGTTCTAGCTGCATCCTCGTTTTAACCATCCAAATAGGATTCA
TTAAGGTATTTGTGACAAAAGCTGCAGAGCCAGCTGAGAGAATGTGCACAGT
ATTGCTATTAGGCACGAAGATGCCATTGAATTGCTCTTTGGCTTTTGGAATAA
CATGCAAAGTAC (SEQ ID NO:45)

Lined number 214
>c0748
GTACTGTTGCACTTCATCCAGAAGTCCCCGGGTGACTATGTTCCCCAGGAGAA
CATCCTGCAGGCCTTGGGTATCTCTGCAGAGGTTTGCTCCAGCAAGCCACCCC
AGTGTGATGAAGAGGTGTGTGGGAGGTGACACGCACAGCCTGAGGACCCGG
AGGCATCTGTAACCTTGGATGTGTTTGGAACACAAAGAATCCCTGTGGAGCC
ACTGAGCCAGGACGTTGGACCTTGTTCATCTTTCAACAGCAATGAGCCATTAA
AGTGCAGGTCCTGGGCAAA (SEQ ID NO:46)

Line number 204
>c0762
GTACCCAGGCCCCTTTGCAAGATTTAATCTTGATTATCTCCTTTTATGTTAGAG
GAGATATAGTATTTAGAACACTCTAAGATATGTTTTATGTTCCTGAAGCTATC
TTGCTCATGTTTAAGATACTAATTGTTTAAATGTAAGTTTTCAATAAACGTAT
TTCTTTAGAGCTGCC (SEQ ID NO:47)

Figure 2L

Line number 185
>C0784
GTCTGAAATATGTCTGAAAGAAGCATTGGAATTGTAAAATGCCCTATGAATT
TGCCAATCCTTAGGTGGAACCAGATAGCCTTTGATGCATATAAAGGACAGCG
GATCCGCTGCACTCCAGGTGCTGCTCAAGCCTCCTAATCAGGTCTAATTGAAC
TGCTGTGGACAGAGTGTGACGTCTGCTCAGGAGGAGAGGGTTTCACTGACCT
GGCTTTTAACAGTTCAAAGCAACAGGAGCAAAACACCCAAGTTGCGCATGAA
TCTGGCATGGTGTCTGATAAATCCTTGTGTGTGCCCCAAATGTGAAGGTTATG
GGCTATTCATTGTTTAGCCAGCTCTTGACATTCACCTTGGGAGTATGTTTGGT
AATCACTTGGAGACGCAGGGGAATAGTCAGTAATGTAGCATTTCTGTTGTAA
TTTCAATTTAACTATCTAGTAGTAC (SEQ ID NO:48)

Line number 227
>C0785
CCCCACTGAAGGCTGAGTGAAGGGCAAAAGGTGAAGGCCTCTTGAACAGCC
CCAGCTGATGCCCATTTTGCAGCAGTTATTTTTTGATTAATTTTGGCAGCCTAT
GGCCATTGTTATGACGACGACTTGAAGAACCTGACTAAGTATAACCTGACTA
AGTGTTCCCTCCAAAGGGGCATGTATCTTTCCATGGGGCAGGCCCTGCCCTGT
CATGCCAGCCAGCGTGCACCTTGACACTGGTGCCGTGGGCAACACAGACATT
CAAGCCTACTCAATATGACAATTCCTCAGAAATGAGGACACTTAGCTGACAC
ACAAAGCACCTTTTTATACAGATCAATGTTTTCTCATTAAACTGACTCTATTC
(SEQ ID NO:49)

Line number 207
>c0793
CTAGTTTATAAATAGAAGTTTCCAGCATCCTATTTGGAAGATTTTCTTTTCTAA
TACTTCCAGGCATTGGATGATTTTAAAGATCTGTTGGTAGTGACAGTAAGGTT
AGATGCTGTTCAAGACAACCAAGTGACTTGAGCAGAGACAACCAGGAAACG
GGTATTCACTAAATACTCTCTTCCTCACCTCAGGTTTGAAGGGAGGATCTGCT
TCAGGTGTAGTCTGGAAGGTTCCTTGGTATGGCTCATGAACACAGAGCTCAA
AGATACCGTCTTGAGAAATCCTCCTTGGTAC (SEQ ID NO:50)

Line number 246
>c0794
GTACCTTTGGAAGAGCAGTCAGAGCTCTTACCTGCTGAGCCATCTTGCCAGCC
CAGTTCTTTTTCTTTTTAAGCATCTTTAATCCCAGCTTCACTTTCTCTGTCAATC
ATCCTTAGGCCTTATACAGATCTTAGTGGAATCCACTCTATCAACCACATGGA
GAGAGAAAGCTCTAAGAGCAGATGTAATTGGCTATGTTCAGAGCAGTCTGCT
CATGGCAGCTGTAGGACACAGAGAATGGCTATATTTCCTCCTTTGTAC (SEQ
ID NO:51)

Figure 2M

Line number 247
>c0823
GAGTCCGGGCGATATAGATCCAAGGCTTGCCTAAAACTCTGTTTTGCTGAAA
ATGATCTTGAACTCCAATCATTCTGTCATAGCTCCCAAGTGCTTGGATGATAG
CTTGTGTCACCATGCTTGGCTGTGTGGGTTTCTGTGGTCTGGGTGGGTTGTTTG
GTTGGTTCTTCAATTGGTTGATTTAATGTAGCCCTGGCTAACCTGGAACTCAG
TATATGACCAGGTTGACCTTGAGCTAGTGGCAATTCTCCTGCCTCTGTCTGTG
AGTGCTGATGCTTTTGTGAATACCAGCTGCCTGCTAGGCAGAAAGAGAATGT
TAACTGAGGTTGTTGCTTGGAGTTTTACATTTTTAACTTTTGGTAC (SEQ ID
NO:52)

Line number 256
>c0840
GTACCCCAGGTGAGCATAACCTTGGGGGCCAACAGTCTGAATGTATCTTTCTC
CCCATTTAAACCTGAGCTGCCTAATGCACAGTTGGGTAAGGGGTGGGGACCC
AGGGCTGAATTGTGGATCCCAGGGAAAGTGCTATGCACAGGAGTGATTTGTC
ACCTCTGGAGCCTGGCTCCTCCTTGGTCTTTGCCATATAAGCACTTATACAAA
GAAGCTCTGGGGTGTTGGGGCTCCTACTGCCTTGGCTGCCCTCTCTCTCCCCT
CCCTGCCTGTGTCGCTTCTCTTTGTCTGCTGTCCCTGAAGGAGATGGCCACCC
TTTGAGCTTAAGGGAGCTGTTTTAGGCAGAGCCTCTGCTTGGGGCTTTTCTGG
AGCTGCTGATGGATAGTATCTGTCCAGGGACCTTGTTTCAAAGGGAATGGGC
ATGGGCCAGCAAGGCACCCCCACAGCTATCAAGAACGTTTTCTTGTTTTTAAA
CCATCACGTCTTCATTTCACATTGGAATAAAGTGAGTTTTTGCAAA (SEQ ID
NO:53)

Line number 215
>C0846
GTACTTGGGAAGCTGAGGCAAGAAAACCAAGGTCGAGCCTAGCCCAGGCTA
CATAGTAAGAAGCTGCCCCTCCCCCCACCCCTTCCAAATCAAAGTCTCAGGA
TGCTCGTTAGAAGAGCGCGCTTTCCAAGCATGCAGAAGGCTATTTTTTCCTCC
CTCACAGCAAATTATTACATTATTACTGGTTTAATTAATTGTAAAGAGAACTT
AATGAAACCACAATGGACACCACGGACTTACAGGTGATTCTCGTGTTCTGCA
CCTGGCTCCCAGTCACATGCCCATGAGTAAGCAATCTAAATCAGCAGTTCAA
AGTGATACGTGACTATCAGCACGGGCGAAGTCCACTAAATCTCCTAGAACTC
TTTAAGATGAAGGAACTATGAGAAGGTATTTTAAAACACCTCAGAGCACTAA
AAAGATCACATTTTAAAACAACATAATCAAAACTTCACATTTTTCCCTGATAT
TCTTAACTAAAAATCTTCTTGTTTTTCATAAATATATTCACAAGAAC (SEQ
ID NO:54)

Figure 2N

Line number 202
>c0859
GTACCCGCGCCATGGGCCTCAGCTCCAACTCCGCTGTGCGAGTTGAGTGGAT
CGCGGCCGTCACCTTTGCTGCTGGCACAGCCGCTCTCGGTTACCTGGCTTACA
AGAAGTTCTACGCTAAAGAGAATCGCACCAAAGCTATGGTGAATCTTCAGAT
CCAGAAAGACAACCCGAAGGTGGTGCATGCCTTCGACATGGAGGATCTGGGG
GATAAGGCCGTGTA (SEQ ID NO:55)

Line number 217
>c0884
GTACCACTAACAGCTATTTTTAACCAAAGTTACTCCAGAGAATTAGAGAAA
TATGAAGTTGTCTGAAATAATTTATTGAAAAGTAAGATTTTTAAAATCCAATC
TATCCTGGTCTTACATTTAATTAAAAACCAAATGATGAAACACCATTTCAGAC
TTATATACATCATGATCCAATTCTTAGAAAAATGTAC (SEQ ID NO:56)

Line number 216
>C0901
GTACAGATGAGGTAATTTACAACACAGAAGTTGTTACTCTGTAAACCCTGAC
CCTCCCTACCCCCCACCCACTCAGATCTTTTAAAACTCTCCACATTTCTGCAA
TGACTCCTTTTT (SEQ ID NO:57)

Line number 220
>C0935
GTACAATATTAACAGTTAGAGAACTGTTATTCAGAAAAGGTGTTAAACATGA
ACAGCAGACTCGGGCTGACTGTCAGCTGTGGAAACATAAGAGGTAACATCAG
TTTAAGTGAAGGCGAAGCCTTGGTCCACAGGCAAGACTCACACGAGGACCAC
AGGACTCAAGAACTGGGAAACTTACGGAGTGTGTGGGCACTGGTCAGTGAGT
ATGGGGCTTGTTCAAAGTTTATCTCCTCATCTATAAAATTATTATTGACATGA
TCTAAAACAAGACCAACTGGGAATGATTGTCATCAAAATCTAGAAAATTCTA
TTAGGAACAAACTATGGAGCCCAAAGTTAATTGAAGAATGGATACTTTCCTA
GGCAGAGTTTTCAAGTGTATTTTCCAAGCAACATCACACAGACTCATAGGCA
ATGATGCAATTTTTAAATAGACAAGATTTTTTTCCCCCTCAATACCTCAGAAC
TTCATAGACATTGTGTTGGAGAATCTGGTCACAAAA (SEQ ID NO:58)

Line number 239
>c0943
GTCAAGACCTAGGACCTAGCATCCCAATTTCAACCTGCACCCCTCATTACATA
AGACTTGTTTTAAACCACGCCGATTACCCACTAATTGGCTCTAAATGGGTCAT
GTGCACTTGTGGATTATCTAACAAGTGGAGACACAGAAGAAACCCTGGTGCA
GGCCAGGCCGGGAGCAGGGACAGTGTTGGCAAGCCAGCTTGTGAGTGTCAG
ATGCTTGGGCACCACGGATGTGAAAGGTGCGCCTGGTGCAATGTATGTGTTT
GGTTAAAGAAACTCTCTGAAATTACTGTTATAATAAGTTTTTAAAAGATTTTT
CTTTCTTTTTAATTTTCACCTTAACTCTTAAATAGGGTAATTTCAATGACCTAG
ACTCTTAGAAAAATTTGACTTACCCCACAACTGACATGTTTCTTTTAGAGCTT
TTGTAAACACAAAATCCTAGTGTAACTT (SEQ ID NO:59)

Figure 2O

Line number 201
>c0974
GTACTAGTTTATAAATAGAAGTTTCCAGCATCCTATTTGGAAGATTTTCTTTTC
TAATACTTCCAGGCATTGGATGATTTTAAAGATCTGTTGGTAGTGACAGTAAG
GTTAGATGCTGTTCAAGACAACCAAGTGACTTGAGCAGAGACAACCAGGAAA
CGGGTATTCACTAAATACTCTCTTCCTCACCTCAGGTTTGAAGGGAGGATCTG
CTTCAGGTGTAGTCTGGAAGGTTCCTTGGTATGGCTCATGAACACAGAGCTCA
AAGATACCGTCTTGAGAAATCCTCCTTGGTAC (SEQ ID NO:60)

Line number 235
>C0983
GGAAAACTATTTTATGCAAAAAAAAAAAAAAAAAAACTATTTTCATCAGTTA
TTGCCTCAGATACGAGAGAAACTATTTCTGGATGCTGTATTGTCAGGTTTTTA
AGTGGGTATTTTTCTTTGAAAAAGAAGCTAAGTGTCTTCACCCACTAGGGACT
TGAAGAAAATTTGAAATAACCCCATATACCAATGATTTCTATGTTAAAGAGC
CTGTTTACACATCAGAAAGCCGATTTTGTATGCATATGAAGGGCCCTTTGGTT
ATGAGCAGGGAGCCATCCAGATGGTCTTTGGTTACTATTTATTGGGGAACGA
TCTGACACTATTACTCTTTTATTTAAGAAACAACGTTGATATGATCAAAGTAT
GACCACTAGAGAGGTTGAGGGGACATGCTTAAAGAGCATAAGCTGACGCCAT
GAAGGGAACACCATACTAGACTTTTTT (SEQ ID NO:61)

Line number 228
>c0991
GTACAGAGAAGCGTGTGCGACCAACCTGCAGTTACAGCTTTCTTCACATTTCC
CCATTCCACACAGCTGTTCCAATGAGCGATCTTGCCCTGGTCACACCACTTTT
AAGCACACTCACAAAACAACCACATATACAGGGTTTCACAAATATGCAAAT
AAACAGGGTTTTCAGAACCAATTTTTACAAAGTGTTATCTAGTAAGATGACCC
AGGGCCCAGTCTCTTTTATTTCCCAAACTGCTTTGACTGACAGTTTAACAGCT
TCCCCTCTGCTCAGGGATGGGTTTTAAGCACCATTCAAGCCAAACCCACAAT
GGATGGATTCAGTGGCAAAGCACTCATGAGGGTTCTACAAGCATTACAATGA
TCTCATAGCAAGCGCAAGACCCTGTCCCCTTGTCACAATGACTGGTTTGAAAG
GGTAAATAAACCAACCCTGGTGTGTTTTGGGCCATTCCAGAAAATCAAGATG
TGTTTTGAA (SEQ ID NO:62)

Figure 2P

Line number 219
>C0996
GTACAGAGTCAAATCACAAAACAAAGGGTAAACGTAACTGCTCAGTCCCTGA
GCCCCAGCACACACACGCACACATACATGCTTACACTTGCTCACAGATTCAC
AGGCAGTCAGCTACAGATGTTAAAGTCAAATAAAGATTTTTTCTTTTCCTGTC
AGGTCCTGGGAAGAGGGGAAAAATGCTCACTGAGAGACACTGGGGCTTCTCA
ATGCCACTCTCTGCTCTGCCAACAACCGAAAAGACAGAAACACTGGGGAGTT
AGCTGTTTCCTTGGCACAGAACTGAATAGGCAACAGAAGGTTAAAGATTATG
TTCACCTTGGGTTTAGTCTCTTTTAACACCACAACAGGGTAGGTAGCAAGCCA
GAAAAGAAAATGTCAGCTCTATTAATTCCAGGAAAAAGAGAGAGAAGGATG
TGAAAACTTGATCCCTATTCACCA (SEQ ID NO:63)

Line number 248
>c1023
TTGTTTGAGGATTTTATTTCAAGACAGCATCCAGAATACGATTCATTTAAAAG
AGCAAAATTAAGTATTTTGCCTCATTTTACCCTTCAAAAACTCGTAGCAATGC
AAATGCAAACCGCAAACCTCCCAAACGTGCCTCTCTGTATAGCTTTTGGAAAT
GAGAGTTCATTGAAAATATTTTTCCATTTTCAGATGAGATGAGATTTGTTGGG
AAATCTTTGTTACAACTTGGGCATCTTGTGGGGAGATAAACTCTCGGACTGTC
CCAGAAACTCCTGGATTGATCTCATCTTCCTGGCTTCCTGAGCCCTGGAAGAA
CAATCCAAGAAGACCTGGATCTTTTTCCTGCACGCCTTCGTCGCGGAATTCAT
TCTGCTTCCAGCAAGCCATCATCATCGACATCTTC (SEQ ID NO:64)

Line number 218
>C1032
GTACTTTTGATGGCTTAGCTCACATTAAAAGTCTGATTAGATTTTTCACCAAA
TCTTCTCTTCAAAGATCACATTCTCATAGTAGTCATAATCTGAAAATAGATTT
TTTTTCCTTGTCCTCAAAGATATCTCTTTAAGCCCACCACTACTAGAATTACAC
CTAGGGCACTAAGAATCGCACATATGAATTATTGCATAGTGTGCAAGGTCTA
GGAATATGATTTCTATCTTCAGGCACTACCTTCGATACCCTGCCTTCTCAAAA
AAGGTGGCCACTAGAAGGTGATTCATCAGGTAAAATGCTTTCCCTGCAACCC
TTACAGCCTAGATTTGATCTCTGTGACCCATAGAAGGAGGAAAGAACCAGAG
CTGTGTTCAGCATGTGAGCCATGGCATGGTGAAGCTCATACATACCACAAAC
AATAAAAGACTTTTTAAATCACTGGC (SEQ ID NO:65)

Figure 2Q

Line number 237
>C1071
TGGGGACGGCGGCTGCCCTCACCTATGGCCTTTACTGCTTCCACCGCGGCCAG
AGCCACCGATCCCAACTCATGATGCGCACCCGGATCGCTGCACAGGGTTTCA
CGGTCGTAGCCATCTTGTGGGGTTAGCGGCATCTGCTATGAAGTCTCAAGCCT
GAGTATAGCCGGGTCTTAAAGCGCCATGGAAACCATTACAAAACCCAGGAAC
AACAGACATCCCTGTCAGACTTGCTCCCTCCGTTTTAGACCGGACCTTATTGT
CATTTGGGTGAGGAAGTGGGCCGATTTTGTAACTGATTTGCGCTTTCACCGGT
GCCCCCTCCCGCTCCCAAAATCCCAAGGTTTCATTTCAGTTGGGGTTGCATGC
TTCTATTTGTGGAAGGCGTCCCTTAATTTACTTTAATAAAAAGCTTATTACAA
CTTGGTATTGT (SEQ ID NO:66)

Line number 221
>C1099
GTACCAAAGGTCAACTGCAACTCTGCAGGGCAGAAAAGCCTGTCTTCATGCC
TTTTCTCCCCTCACATGGAGGGGTGGCTAAGTCTGCAACACTCTAGAGGAAA
CAATAATGTGTTTAAGATTGTCCTGTAGGTGTCTTCCTCCCTCTCCCTGGCATC
CCAACCACCGACATAGGTGGGTGTATAGTAAGCGTCTTTCAGAGATGACATT
TTGCTATTCTCCTTTTGGGGGGAAAAAAAAATCAAGCAGAGAAAATTTGATA
AACTCCACAGTTTTAGTGTTTGTGAAAAACAAGCTGTTCCTTATTTTGTTCTCA
AACTCTTATTTTTCCCAATAAGTCATCATGTGTTGCTGACAACAACTG (SEQ
ID NO:67)

Line number 242
>c1112
GTACATATGCACCAAATTCCATTTTAGAAGTTTCCATATCATTTTCATAGAAA
ACAAAGTTTGAAAACAAGTTAACATTTAAACACAGCACGGTATTCTATCACA
ACTGAAACTTTTCTTCTTTACAGGACTCAACAAAATCTAAAAATGAACTATGC
TGTAGATTTACTTCATGCAAAGATCTTTATGTTATCTCTGAAAATGAAAAGAA
TGGCTTTAAAAGCACATTTTATACTATTATGGCAACTTGTGTAC
(SEQ ID NO:68)

Line number 222
>c1144
AGACAGTGCTGCCTCCCTGGTGGATACGGTGGGATCCAGCTAGGTCATCTGA
AAGATTAAGTCCCAGGATCTCAGCCAACAGTGACCATGCTTCCAGGGAAACT
GCTTCCTGCACTTGCAGAACCACCAGTAGGCTGCAGGTGACTAACGTATCTG
CAAGGTTAGAGATGCTCGGTGTTTCATACCCTGCGTTGGTCCTAGAGAGAAG
GGCCTTTGGCATTGCTTTGGGCAGGTAGTGGGAGAAGAAAGGGATCCTGGGT
GGGGTGTATTTATATATAATTTCGGTTTTGTTTGTAC (SEQ ID NO:69)

Figure 2R

Line number 223
>C1148
GTACAGATTATCCCTCCTGAGACGCTGTGCCATCGGCCAGCACTCTCACTCTT
CCCACACCTTCCCTCCTGTCCTGCTACCCACTGTTAGGCTCTCAGTGTCTCAG
ACATCTGTAAGCTGCTACTTCATCTGCAAAGAAAATAGCTGTTCTTTTAAAAA
TTAAACTATGTAC (SEQ ID NO:70)

Line number 224
>C1176
GTACTTTTTGTTCAGTCACAGAGACATGCCACCAGCCTGTTCAGACCAACTGA
GTCAGGAGAAACATACAACATGTGAAGACTATCTCTCTCCAGCTGGTGACAC
AAAACAGCCCAGGTTCCAACCACAGGATTATTAACAATAAAAACTACACCGC
TAGCCTGGCACGGTGGCCCGGACCCATAATCCCAGCACTTTGGAAGCTGAGG
TAGAAGGTCTGTGGCCAGCCTGGGCTACAGAAGTGAGACTCCTGTTGCAAAA
AGGAAAACAATACACACAAATTGTAACCCTGCTTTTTCTAAGGAAAAAGTAG
TTTTTAAAACAAGGGGCAAGAAATAAACTTCAAACTGAGACTCATATT ((SEQ
ID NO:71)

Line number 225
>C1201
GTACAGTGATCTAATGTTTAGGAGGCCTTCCTTTGACTTTTTATATTCGAGGC
AACATCTTTTTTTAAAAATCATGAATTATAGTTTTTTAGCAAGTGACCGTTAT
GTTCTGTTTTCCTGTCGGGATCGTTTGGAAGGAAGGAGAAAACTGGCTTTGCA
GGTGTTACTGGCTGGCACAGTAC (SEQ ID NO:72)

Line number 226
>C1205
GTACTGGACTTTGTTACAAAGCGACCCAGTTGACTGCTTTGACAGGAGCAGA
TGAGAAGCATATCTTATGTTCTGTAAGAATGAGGCTCTGAATCCCTCTCTATA
GAAAAGGTTACAGGCACTGTTCTTTATAACCAAAGAGCTTCCAGGAGACTGA
ACGTCTTTGCTGTCCACTCTTGTTTTGTGTGAGTATATAATCTCACTCCAATCT
GGTGCCATACTTCCCTTGGCTACCAAGCCACGTGCTGCCCTTGGTCCCTGTCC
CCTTTCCTAAGCACACTGAGAAATCGCACAGCTGTAACCTTCAGTCTTCCACA
TAGCCTGGTAC (SEQ ID NO:73)

Line number 206
>c1217
AAACAATTATACTCCTGAATTCTCTTCTCTGATATCTTTTAATATTTTAATGTG
CACCTTTTTCTCTATATAGACACACTCATATACAGTTGAAAAGTTAAGTCAGA
CTCTTCTTTATAGCCTTTCTGCCACTTGTTTCCTGACACCAGCTTTTATGAAAC
AGACATATTTTCAAGAGTTTTACTTCATTTTTATTAATAATACCACAGTATTCT
ATCATTTACTATATTAGGATTCATTCAGTATTGAATATTTAGATTGTTTTAATT
TTTACTATCTAAAAATATGTCAGACTCATCCTTGGGTGCAACCTAGTATGCTT
G (SEQ ID NO:74)

Figure 3A
Adipocyte Subtractive Library Sequences

Line Number 55
>b0010
ACCTTGCTCTGCATGCCTTTCTCCCCCGTGCTGTCCCCTTTGGAGAGGTCACC
AGGAACGCTTGAGGCAGTCTCCTGTCACTTTGGAGTGTGTATGTGTAGCAGGT
CCAACAGGAAAGAATAAGTATGGTGTTGGGCTCAGCAGTCAAGAGCACACCC
TGCTCTTACAGAGGACCTGCGTGTGGTTCCAAACACCCACTTCGGGTGGCCA
CAAACCTCCTGTAATTCCAACCCCA (SEQ ID NO:75)

Line Number 47
>b0037
ACTTCCCTCCACTTGGGGACCACTCCACTATGTGAGCATTTTCCTTTATATTTT
TTATGAAAGCTGATCTTCCCTCTATAAGATTCCATGTTCTTAATGTTTTATCGG
TGCCAACAGACAGAGCCAATTTGCCAGACGGGTGGATAGAAAGGAAAGTGA
CATGTCCCCTGTGTAGCTTTAAAGGTCTTCAGGCACTTCCATCTCTTCACATCC
C (SEQ ID NO:76)

Line Number 59
>B0065
ACATGTTAATGATTTATCAGTATGCCCCGAATTCCTAGTGCAGTTCTCATTCT
CCGCATGCCCTCAGCTGTGGTCAGGTGACTTCCTGTCCCCTGGCAGCTCTGCT
GAGTCCCTGTGTTTGAGCCTCCAGGGAGAAGGGTTTGGGCTGCATTCTTCTTA
TCCCCATGCACAGAAACGCTCAGGGTCCCCACGTGCCTGTTGTCCTCCCCTCT
AGTCCTTTGTTCCCTTTCTGAGCATGTGGCCCCTCCCCCGGCTGTGAGGGGCA
CCCCTTCCTAAGAAAGTGTTAATGCAGGATCAAAGACCACCACTGCATGTGT
AC (SEQ ID NO:77)

Line Number 57
>b0091
ACTTCCTTGTGTCAAAGCAAGAACATTTCCCCACGTTAACCCAAACCTCCACC
TCCTATCCTGAAACCCTGACCACAGGACTCACGGGAGTGTGCCGACAGCAGG
TGCTATGATCACAAGATAATCAGAGATCCAGGAAGGAAAGGGAAGGGAGGC
TGAAGGAACAGGAAGCATGGGATAAGCAAGACAAAGTGTGATAGCTTTCAA
AGAAAATGTTGAGTTTCTTTTGGGATGGCTATGCTGTTCTTCA (SEQ ID
NO:78)

Line Number 52
>B0096
TGGGAATAAATCAAAAATTGATCAGTCTGGTGCAGGAGCCAGGGAGCTCTGG
CCTGCACATCCTTCTGTGCCATCTTAGATGGCAGCTACCATCAAGGTCAGGCA
AAAAGCCTTAAGACAGCTCACTGAAAGTCTCGTGTAAGGGCATTTGGTTGGC
GGTTTTCAAGGGCTCTCAGAACACCAGCTCCACGTGCAGGGGAAGACCATTT
TCCAAATAAGCCATGAATGGGCCCTCTCATCTTACCCAGTCTCAAGCCCTATT
TT (SEQ ID NO:79)

Figure 3B

Line Number 64
>B0101
ACCATATTTCAAGAAGATAGCAGAAATTCAATGAAGTCGATTTAAAAACCTG
GGGCTATTTACCTGGTCTGAGGTATGCAAGCTGTCCAAAGGTGGACACGTCT
AAATAAATTTACCATGTCAGTTATCGGCCAAATGTTTTAGCATTTTTTTAAGC
AGGATTTTCTTGGCTGAGAAGACAAGAAGACATTTTAATATCCATTGTAGAA
ACGAGAATGGGCTGGGCACTGATGGTGCACGCCTTTAATCCCAACACTTGGA
AGGCAGAGGCAGGTGGATCTCTGTGAGTTGGAGTCTAGCCTGGTCTACAAAG
TGAGTTCCAGAATAGCCAGGGCTACACAGAGAAACCCTGTGTTGGCCAAACC
AAAGC (SEQ ID NO:80)

Line Number 63
>b0117
GATACTTCATATCTGTCTTCAAATATGTATTTCCTAGGTATCTTTTCCATTGTA
ATTTTCAGTAAATATACTGAAGCTTGATTTCACATTCATTTCTAAACTTGAAC
TGTTTTTAATTTTAAAAATAGGCACTTATACAGAGAAACTTATAAAAACAATT
TCATACATACTGAGTTGTCTCAGCATTTCTTTAAAGTCATTTGAGCTTTCTAGT
TTTGCTGAAACTGCATTCAATGAATCATTCAGAAAAACAAAACATATCAGAA
GGAAAACTATCTTTTTAAAATTTTGTTCTATAACCCACCTTTTAAAAACTTTAT
ACAAGTCTATATAATACTCTTTTTATTGTCTAGGAAGAAATTACCCTTCTACC
ATTAGCTTTAACAT (SEQ ID NO:81)

Line Number 53
> b0129
ACGAAGAGTCAGACAAAATTCCCCTCTAGCTCACCCATCCGTTGCTGAAATC
TCTTTGGGAAGGTGGGGGAGATATAAGCCACATCTGCCCCTTCTCCTAAGGTC
TCACTGACAAACTGAGGCAGGGTTCTGTCAGACTTTACTCCAAGGAACCAGT
GAGTGTCATTAAAGCAGAGTGGTGTATAACAGGTGGTAGGGCGCCACTGGAT
ACACAGATGAGAGTCTCCTGGTCCTTGACACCACTCCACAATCAGATAAAAT
CAGTCAGCAAGGCCAGCTGGGACAACCCTTTTAAAAGGGGGACAGCTGACA
CCTCCACCCCCCTCTGGGCTGGTGGTTGCTTGCTTCTGAGACAGTCACTCAAA
TACTACCCCA (SEQ ID NO:82)

Line Number 51
>b0136
ACATTAATGATTTATTAAAAGAAACAACTCCTTGTCCCACTCCACTGTGCTGC
TTGTAATCTCCATACATGGCCTCCATTTTCAACTGTTTTCTTGGTCACAGAACT
CCAAACAAACACATTTTTTTTTCCAGGTAAAAGCTGTTTTTAGTTTGTAGT
(SEQ ID NO:83)

Figure 3C

Line Number 54
>b0154
ACCTATGTCATGGAAGGGAACTCTGTTACAAGCTTGAGCATATTTCAAATAA
CTGTGGTAGTGCTGAGAAATAGGCTTTGCATCTGGCTAAGAGAAAAGAGCCT
TCATGGAGTATCACAGAAAACTTCCAATTCTAGGTTTTCTATATACTTTGTTA
CCTTGTTCAAAGAGATTTAATTGCTTCTATAATCAAAGAAGAGAACATGGAA
CCCCCTCACTTCTAGGAAGTAGATGTGGTAACATTATTCCAGAGAATCTGTCA
CCTAAAACTGAAGTGTTTACAATCCCTTAAAGCTAACCGGCTCCCCAGTGTGT
TATACAAGGCAAGCCAAGATAGCCTCACTGTCCCTGAACTGGAAGTGTCACA
CAGCTCAG (SEQ ID NO:84)

Line Number 26
>b0158
ATACTCTTTAAATATATTTTCAGTTTCTGATTTAGAGACATTGTTCACCACTGG
ACTAGACTAACCTTGGATGCCGGTCAGTGACGACTTTGTCACACTAAATAAG
ATGGCTGTTCTAGGCCTAAGACACTGTCTTGAGGGTCAGTAAGTGACGAGTC
AGATGAGTTAACTGTGGGGAAACTAGAACATTCGAGCCAAGAATTATTG
(SEQ ID NO:85)

Line Number 56
>B0174
ACATTTGAATCCTTGTCAAAACAACAAAACATTAATGAGTTACTTTAGAGTAT
CTTTGTTGAAAATGGAATTTTCAATAACTGTTTCATAATGTTTTGTTTTATTAC
TCCTTAAGACTGATACTATAAGATGAAGATAACATTACTTCAAAATAGGTTC
AAAGCATTTATTTTACAACTTTATGTTTTTATAAATTTCAGAATAAAAAAAAA
AAAGCTTG (SEQ ID NO:86)

Line Number 58
>b0175
ACTGTGGCCTCATTCCTCCTACCCTCCAAGGGGTGCGCTATGTGGATGGCGGC
ATTTCAGACAACTTGCCACTTTATGAGCTGAAGAATACCATCACAGTGTCCCC
ATTCTCAGGCGAGAGTGACATCTGCCCTCAGGACAGCTCCACCAACATCCAC
GAGCTTCGCGTCACCAACACCAGCATCCAGTTCAACCTTCGCAATCTCTACCG
CCTCTCGAAGGCTCTCTTCCCGCCAGAGCCCATGGTCCTCCGAGAGATGTGCA
AACAGGGCTACAGAGATGGACTTCGATTCCTTAGGAGGAATGGCCTACTGAA
(SEQ ID NO:87)

Figure 3D

Line Number 48
>b0188
ACCAAACCCCATCAAATACATAAAAATGCCTATCTCCAAACACTGCATGTCT
AGACATTTACCCTGAAGAAATACTAGCATTAGTGTGCAAAAAAAAAACAATG
CATACTAGGTTGCACACAAGGATGAGTCTGACATATTTTTAGATAGTAAAAA
TTAAAACAATCTAAATATTCAATACTGAATGAATCCTAATATAGTAAATGAT
AGAATACTGTGGTATTATTAATAAAAATGAAGTAAAACTCTTGAAAATATGT
CTGTTTCATAAAAGCTGGTGTCAGGAAACAAGTGGCTGAAAGGCTATTAAGA
AGAATCTGACTTAACTTTTCAACTGTATATGAGTGTGTCTATATAGAGAAAAA
GGTGCCCATTT (SEQ ID NO:88)

Line Number 60
>b0217
ACTAATTGCAGTTATAACCCAAAAAGACTTCAAAGACATGAAAGAGACTCTT
GGAGATGACATTACCGTGAAAATGTATTCCCCATCCTGGCCTAACTTTGACTA
TACTCTGGTGGTTATTTTTGTAATTGCTGTGTTCACTGTGGCCTTAGGAGGAT
ACTGGAGTGGACTTATTGAATTGGAAAACATGAAGTCAGTGGAAGACGCCGA
AGACAGAGAGACCAGAAAGGGAGAAGGACGATTACTTTACATTCAGTCCTCN
CACAGTTGTTGTGTTCGTGGTCATCTGCTGTATAATGATTGTCTTACTGTATTT
CTTCTACAAGTGGCTTGTGTTTGTTATGATAGCGATTTTCTG (SEQ ID NO:89)

Line Number 62
>B0237
TTTTTGATCTTGTTTCCATTTATTAGTCTAGGAATATACAAAATTGTGACATGG
CACAACAATCTCATGGAGAAAACAAAATACACAGGAGGGTAGCTTTCTCCCA
ACCCCTCATAAAGGCACACACTTGAGGCTTGGTTTGGGGGGGTCCTCCTGTCC
TGGACCCAGTGTGCTGCACAAAAAGCCAGGCACCTCTGTGAAAAGAAAAGA
AGCCCTTCCTCACACCAAACTTGGGGTGTGGCTCAGGAGGATTTGCCCCAAA
ACAACGCTAAACTGTGTCCATGGTGCCAACC (SEQ ID NO:90)

Line Number 50
>B0245
GATACAAGCAAACTGACTTCTGAAATGGACTATCACACTCTCCACCTGCCGG
GCCCTCTCAAGCTGAGGTGGCTTTTTGCATTTTGCTACTCCTGGAGGCCATAG
GCCAATGGATAATCTATGTTTCTCACTGTCTGTGTTTCCACCACGTTCCTAGCT
CCCTGGAGT (SEQ ID NO:91)

Line Number 61
>b0274
TTTTAATGCAAGTTTTTATTGAGCTGTATATACAATTTAAGCTATTAAAATTTA
TACAATATTTACAAATTAAATAATCATCTGAAACTCTCCAGAAACTCCTGTAA
CACAAAACCAAAAGAAGGAAGCAGCAGCTAGGCCTTGGCTCCCTAGGCCTG
GCCTGGCCCGGCCCGGCCCAAAGCTTGTTCTAAACTGCGAAGTCCAGGT
(SEQ ID NO:92)

Figure 3E

Line Number 49
>b0284
ACTTT.GCTTTTTATTATTCATTCAATAATATAAATATTGTGTCAGTTTCTACAG
TCTCTTGACTTTAATGTATAGCTATCTATAGATCTGCTTCTCCTTATTTATGCT
AGCTTTCAGGTTACTTCTGGCAACTCTGGCTTGCTTGCTGGTGAAGAGGCCCT
GATTCCAACAAAAGCAGAAAGCAGCATTAGATTTTATTTTCCTTGTTCTAAAA
TGCTTGGCTTTTGCTAGTCTAAAAATGATTGATTTTGATTAGAAATTTGAAAA
TGACAAGGATTTCTGAGTGTATCTAAAACTAAACAATTTACCAATGTCATTCC
ATTTACTAGTGTTAATACTGAAGACCTTAACAATTCCGTCCACAAAAGGTAAC
TACAGGGGAA (SEQ ID NO:93)

Figure 4

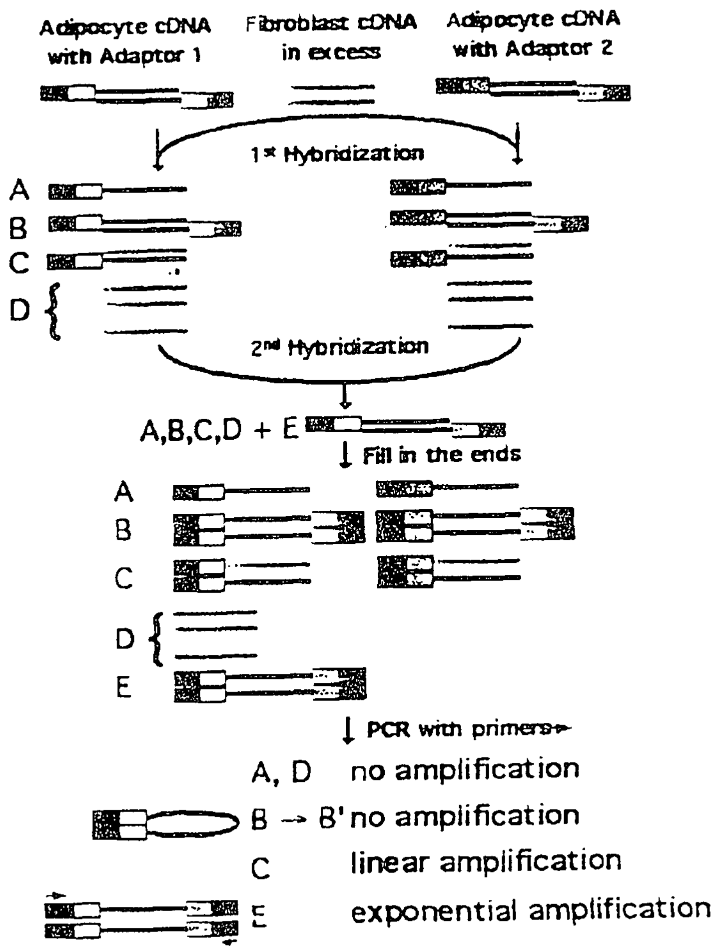
Adipocyte cDNA with Adaptor 1
Fibroblast cDNA in excess
Adipocyte cDNA with Adaptor 2
1st Hybridization
A
B
C
D
2nd Hybridization
A,B,C,D + E
Fill in the ends
A
B
C
D
E
PCR with primers
A, D no amplification
B → B' no amplification
C linear amplification
E exponential amplification

Figure 5

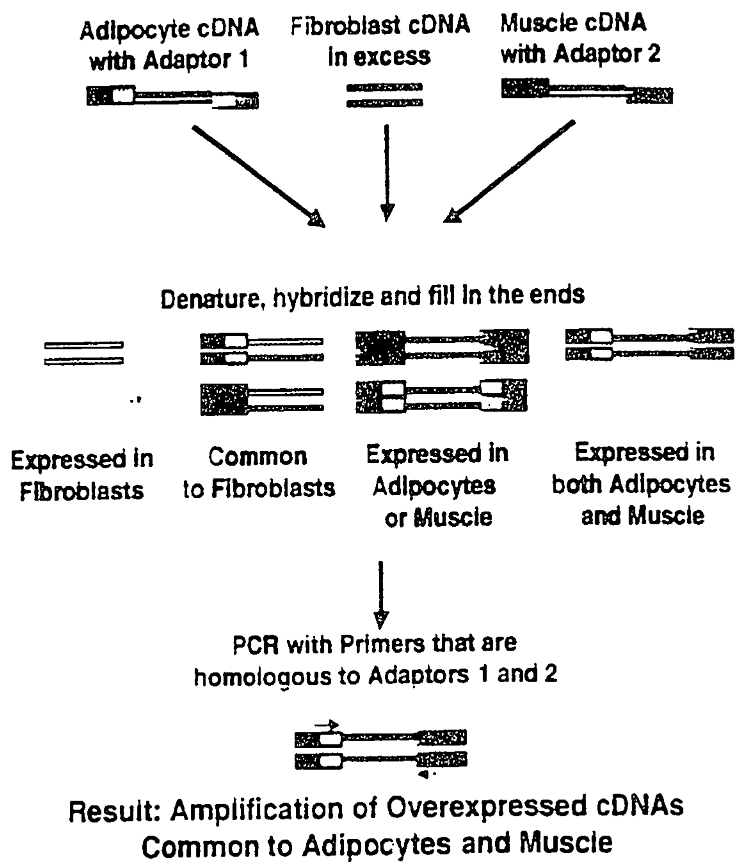
Adipocyte cDNA with Adaptor 1
Fibroblast cDNA in excess
Muscle cDNA with Adaptor 2
Denature, hybridize and fill in the ends
Expressed in Fibroblasts
Common to Fibroblasts
Expressed in Adipocytes or Muscle
Expressed in both Adipocytes and Muscle
PCR with Primers that are homologous to Adaptors 1 and 2
Result: Amplification of Overexpressed cDNAs Common to Adipocytes and Muscle

Figure 6A

Adipocyte Subtractive Library

NUMBER	Name	Accession number	Accession species	Human Accession	Clone numbers	Hits	Class
1	stearoyl-CoA desaturase	M21285 J04190	mouse	AF097514	4 13 20 29 46 50 51 52 54 64 72 76 105 131 120 145 203 207 219 226 230 240 248 250 254 261 287 290	28	genes overexpressed in adipocytes
2	adipoQ	U49915	mouse	NM_004797	6 9 43 44a 45 63 66 69 124 132 133 134 144a 146b 182 201 235 238 251 281	20	genes overexpressed in adipocytes
3	CD36 (LDL receptor)	L23108	mouse	M98399	2 3 5 23 42 42b 68 135 170 171 199 200 224 288	14	genes overexpressed in adipocytes
4	adipocyte P2; lipid binding protein	M13264	mouse	J02874	14 24 62 70 78 80 196 197 198 202 263 273 278	13	genes overexpressed in adipocytes
5	adipsin (possible serine protease)	M11768	mouse	NM_001928	12 18 [illegible] 82 106 125 159 183 187 228 267	11	genes overexpressed in adipocytes
6	haptoglobin	S67972	mouse	NM_005143	1 109 144b 147 148 150 151 280	8	genes overexpressed in adipocytes
7	long-chain acyl-CoA synthetase	D90109	rat	NM_001995	108 118 119 186 222 236	6	genes overexpressed in adipocytes
8	CAP (3' splice difference???)	AL117472	human	AL117472	88 89	2	genes overexpressed in adipocytes
9	medium-chain acyl-CoA dehydrogenase	U07159	mouse	NM_000016	32 34	2	genes overexpressed in adipocytes
10	acetyl-coenzyme A carboxylase	X53003	rat	none	102	1	genes overexpressed in adipocytes
11	adipocyte hormone-sensitive cyclic AMP phosphodiesterase	Z22867	rat	NM_000922	126	1	genes overexpressed in adipocytes
12	caveolin-1	U07645	mouse	NM_001753 3	229	1	genes overexpressed in adipocytes

Figure 6B

13	pyruvate carboxylase	L09192	mouse	NM_000920	155	1	genes overexpressed in adipocytes
14	vascular endothelial growth	U22372	rat	AF024710	204	1	genes overexpressed in adipocytes
15	pro alpha1 (III) collagen chain (COL3A1)	X57983 S37761	mouse	NM_000090	177	1	genes not overexpressed in adipocytes
16	branched chain aminotransferase (BCATm)	U68417	rat	NM_001190	40 239 243	3	genes of unknown expression (metabolic)
17	NADH:ubiquinone reductase	X61100	human	X61100	15 140 141	3	genes of unknown expression (metabolic)
18	cytosolic 3-hydroxy 3-methylglutaryl coenzyme A synthase	X52625	rat	X66435	79 206	2	genes of unknown expression (metabolic)
19	3-hydroxyisobutyryl-coenzyme A hydrolase	U66669	human	U66669	19	1	genes of unknown expression (metabolic)
20	epoxide hydrolase	Z37107	mouse	NM_001979	33	1	genes of unknown expression (metabolic)
21	Inducible 6-phosphofructo-2-kinase/fructose 2,6-bisphosphatase (IPFK-2)	AF056320	human	AF056320	212	1	genes of unknown expression (metabolic)
22	isovaleryl-CoA dehydrogenase (IVD)	J05031 M26201	rat	NM_002225	81	1	genes of unknown expression (metabolic)
23	liver microsomal carboxylesterase	U10698	rat	NM_012122	216	1	genes of unknown expression (metabolic)
24	ubiquinol cytochrome-c reductase core I	L16842	human	L16842	28	1	genes of unknown expression (metabolic)
25	CD98 heavy chain	U25708	mouse	J02939	95 113	2	genes of unknown expression (signaling)
26	diphosphoinositol polyphosphate phosphohydrolase type 2 alpha	AF191649	human	AF191649	158 209	2	genes of unknown expression (signaling)
27	BKLF (transcription factor)	U36340	mouse	NM_016531	220	1	genes of unknown expression (signaling)

Figure 6C

28	G alpha 11	U37413	mouse	L40630	31	1	genes of unknown expression (signaling)
29	homo sapiens GC-rich sequence DNA-binding factor candidate (GCFC) mRNA, complete cds.	AA543412, AF153208	human	AA543412	115	1	genes of unknown expression (signaling)
31	Max protein	M63903	mouse	M64240	168	1	genes of unknown expression (signaling)
32	N-Cor (RIP13) (transcriptional repressor)	U22016, U35312	mouse	NM_00631	61	1	genes of unknown expression (signaling)
33	protein kinase KID-1	AF057026	rat	M16750	92	1	genes of unknown expression (signaling)
34	protein phosphatase-2A regulatory subunit-beta	M65254	human	M65254	98	1	genes of unknown expression (signaling)
35	Rad51d	AF034955	mouse	NM_002878	253 256 258	3	genes of unknown expression (other)
36	mitochondrial elongation factor Tu	L38995	human	L38995	138 139	2	genes of unknown expression (other)
37	78 kDa glucose-regulated protein grp78	D78645	mouse	AJ271729	153	1	genes of unknown expression (other)
38	cytoplasmic gamma-actin	L21996	mouse	M19283	48	1	genes of unknown expression (other)
39	ERp99 encoding an endoplasmic reticulum transmembrane protein (GRP 94)	J03297	mouse	NM_003299	77	1	genes of unknown expression (other)
40	eukaryotic initiation factor 5 (eIF-5)	L11651	rat	NM_001969	83	1	genes of unknown expression (other)
41	Mus musculus ribosomal protein S11	U93864	mouse	NM_001015	247	1	genes of unknown expression (other)
42	ubiquitin protein ligase Nedd-4	U96635	mouse	D42055	107	1	genes of unknown expression (other)
43	cleavage signal-1 protein (cs-1)	M61199	human	M61199	90	1	genes of unknown expression (unknown function)

Figure 6D

44	Mouse virus-like (VL30) retro-element	M21123	mouse	none	180	1	genes of unknown expression (unknown function)
45	mRNA for KIAA0516 protein Also HSS protein	AB011088	human	AB011088	264	1	genes of unknown expression (unknown function)
46	nuclear autoantigen GS2NA	U17989	human	U17989	85	1	genes of unknown expression (unknown function)
47	unknown protein with WD reapeats	AA870601			37 38 39	3	unknown genes with est
48	similar to drosphophila putative type III alcohol dehydrogenase (T3dh)	U66411			188 189	2	unknown genes with est
49	unknown	AA960368			284 286	2	unknown genes with est
50	unknown??? (homology to 3' end of ets 2??)	AA542151			245 246	2	unknown genes with est
51	unknown member of WAP family has a potential leucine zipper	AA260487			136 142	2	unknown genes with est
52	unknown	AA624008			96	1	unknown genes with est
53	unknown	AA850341			129	1	unknown genes with est
54	unknown	AA623774			154	1	unknown genes with est
55	unknown	AW761764			10	1	unknown genes with est
56	unknown 3' end of mRNA in T Cell gamma locus	AA273445			174	1	unknown genes with est
57	unknown contig has repetitive day	AA274981			91	1	unknown genes with est
58	unknown gene homologous to Homo sapiens clone 24519 (unknown) gs2 like	AF055000			175	1	unknown genes with est
59	*unknown homologous to* unknown human gene FLJ20750	AA472160			65	1	unknown genes with est

Figure 6E

60	unknown homologous to unknown human gene HQ0270 and similar to fos39554	AU079046			217	1	unknown genes with est
61	Unknown homology to NOF1 in wrong direction	U39400			274	1	unknown genes with est
62	unknown or transferrin isoform	AA796808			237	1	unknown genes with est
63	unknown possibly homologous to human AAPT1-like protein (homolgy to choline/ethanolaminephosphotransferase)	AA822381			117	1	unknown genes with est
64	unknown with b1 repetitive element	AA501113			101	1	unknown genes with est
65	mitochondrial genome	V00711	mouse	NC_001807	21 86 116 121 173 271 283	7	Mitochondrial

Figure 7B

12	carnitine palmitoyltransferase 2 (Cpt2)	NM_009949	mouse	NM_000098	563	1	genes overexpressed in both muscle and adipocytes
13	glucose transporter (GLUT4)	NM_009204	mouse	M20747	1140	1	genes overexpressed in both muscle and adipocytes
14	adipocyte P2; lipid binding protein	M13264	mouse	J02874	98 100 149 189 191 207 217 210 249 307 317 333 396 426 444 500 521 527 554 558 571 594 602 608 609 615 660 667 669 670 690 691 703 718 733 753 765 792 803 808 809 834 845 858 869 874 877 881 889 890 894 910 911 934 946 948 951 998 1005 1024 1036 1049 1050	83	genes overexpressed in adipocytes, unknown muscle expression
15	stearoyl-CoA desaturase	M21285 J04190	mouse	AF097514	58 67 71 79 123 201 233 252 344 467 479 508 547 577 581 584 668 671 695 698 716 717 747 759 768 769 770 807 816 867 868 907 924 928 945 961 966 970 971 1021 1053 1096 1126 1127 1142 1149 1174 1192 1204 1220	50	genes overexpressed in adipocytes, unknown muscle expression

Figure 7A

Muscle-Adipocyte Union Library

NUMBER	Name	Accession number	Accession species	Human accession number	Clone numbers	Hits	Class
1	CD36 (LDL receptor)	L23108	mouse	M98399	88 174 451 488 495 499 512 514 592 604 643 656 662 709 781 830 917 926 936 964 1040 1095 1125 1151 1183 1219	26	genes overexpressed in both muscle and adipocytes
2	fructose-16-bisphospate aldolase	Y00516	mouse	NM_000034	51 129 136 316 359 378 623 722 750 815 885 927 1045 1137 1143 1179 1194	17	genes overexpressed in both muscle and adipocytes
3	glycerol-3-phosphate dehydrogenase	M25558	mouse	NM_005276	94 267 543 748 897 977	6	genes overexpressed in both muscle and adipocytes
4	hexokinase II	AJ238540	mouse	AF148513	566 620 642 708 878	5	genes overexpressed in both muscle and adipocytes
5	carbonic anhydrase III	NM_007606	mouse	AJ006473	593 836 1177	3	genes overexpressed in both muscle and adipocytes
6	lipoprotein lipase (LPL)	M60847	mouse	M60847	146 1090 1190	3	genes overexpressed in both muscle and adipocytes
7	medium-chain acyl-CoA dehydrogenase (MCAD)	NM_007382	mouse	M16827	560 1074 1168	3	genes overexpressed in both muscle and adipocytes
6	aps	AF095576	rat	AB000520	311 346	2	genes overexpressed in both muscle and adipocytes
9	diazepam binding inhibitor (Dbi)	NM_007830	mouse	M15887	528 1123	2	genes overexpressed in both muscle and adipocytes
10	glyceraldehyde-3-phosphate dehydrogenase (Gapd)	NM_008084	mouse	NM_002046	773 1155	2	genes overexpressed in both muscle and adipocytes
11	probably the 3' UTR of the human insulin receptor via rat est	AA957207	mouse	X02160	551 1032	2	genes overexpressed in both muscle and adipocytes

Figure 7C

16	adipoQ	U49915	mouse	NM_004797	75 363 376 382 436 447 454 456 510 511 516 548 583 630 719 826 837 866 879 893 903 918 920 946 1028 1041 1043 1053 1079 1086 1162 1198 1221	33	genes overexpressed in adipocytes, unknown muscle expression
17	adipsin	M11768	mouse	NM_001928	214 403 469 497 523 542 598 605 629 637 663 713 720 764 774 922 929 931 995 1018 1022 1038 1114 1116 1121 1156	26	genes overexpressed in adipocytes, unknown muscle expression
18	long-chain acyl-CoA synthetase	D90109 J05439	rat	D10040	74 99 606 697 821 1015 1025 1101	8	genes overexpressed in adipocytes, unknown muscle expression
19	2-oxoglutarate dehydrogenase E1 component	U02971	mouse	NM_002541	694 740 900 1103 1141	5	genes overexpressed in adipocytes, unknown muscle expression
20	FSP27 (homolgy to ced apotoctric proteins and nucleases)	M61737	mouse	none	235b 435 683 1161	4	genes overexpressed in adipocytes, unknown muscle expression
21	spot 14	X95279	mouse	Y08409	525 634 1193	3	genes overexpressed in adipocytes, unknown muscle expression
22	calcium-binding protein Cab45b	U45978	mouse	NM_016547	108 1102	2	genes overexpressed in adipocytes, unknown muscle expression
23	caveolin-1	U07645	mouse	NM_001753	229 448	2	genes overexpressed in adipocytes, unknown muscle expression
24	Cdk4 and Cdk6 inhibitor p18 protein	U19596	mouse	NM_001262	38 835	2	genes overexpressed in adipocytes, unknown muscle expression

Figure 7D

25	3-hydroxyisobutyrate dehydrogenase	J04628	rat	AC007130	202	1	genes overexpressed in adipocytes, unknown muscle expression
26	adipose differentiation related protein (Adipophilin)	NM_007408	mouse	NM_001122	843	1	genes overexpressed in adipocytes, unknown muscle expression
27	CAP (3' splice difference???, same as b088)	AL117472	human	AL117472	1060	1	genes overexpressed in adipocytes, unknown muscle expression
28	Dihydrolipoamide Dehydrogenase	NM_007861	mouse	J03620	672	1	genes overexpressed in adipocytes, unknown muscle expression
29	S3-12	AF064748	mouse	none	721	1	genes overexpressed in adipocytes, unknown muscle expression
30	electron transfer flavoprotein beta subunit	X71129	human	X71129	145 402 275 607 631 923 1020	7	genes overexpressed in muscle, unknown adipocyte expression
31	NIPSNAP2	AJ001261	mouse	AJ001259	30 780 786 930 1034 1046	6	genes overexpressed in muscle, unknown adipocyte expression
32	desmin	X73524	rat	NM_001927	68 519 1135 1208	4	genes overexpressed in muscle, unknown adipocyte expression
33	myosin regulatory light chain (RLC)	X05566	rat	NM_006471	531 875 1145	3	genes overexpressed in muscle, unknown adipocyte expression
34	phosphatase 2A B56-alpha (PP2A)	L42373	human	L42373	776 1006 1138	3	genes overexpressed in muscle, unknown adipocyte expression
35	phosphatase 2C	J04503	rat	S87759	452 453 474*	3	genes overexpressed in muscle, unknown adipocyte expression
36	Sui1 homolog	AF129888	mouse	AF083441	595 1057 1110	3	genes overexpressed in muscle, unknown adipocyte expression
37	adenylosuccinate synthetase 1	NM_007421	mouse	NM_001126	534 655	2	genes overexpressed in muscle, unknown adipocyte expression
38	ANC1 (Ant1) for adenine nucleotide carrier	X74510	mouse	J02966	438 969	2	genes overexpressed in muscle, unknown adipocyte expression
39	mlrq-like protein NADH:ubiquinone oxidoreductase	U59509 U94586	mouse	NM_002489	102 916	2	genes overexpressed in muscle, unknown adipocyte expression

Figure 7E

40	myoglobin	X04417	mouse	NM_005368	318 589	2	genes overexpressed in muscle, unknown adipocyte expression
41	pre-B cell enhancing factor (PBEF)	AF234625	mouse	NM_005746	314 597	2	genes overexpressed in muscle, unknown adipocyte expression
42	S100 alpha (CA binding protein)	S68809	rat	NM_006271	84 1109	2	genes overexpressed in muscle, unknown adipocyte expression
43	ubiquitin specific protease 41 (UBP41)	AAF07565	mouse	NM_004205	175 942	2	genes overexpressed in muscle, unknown adipocyte expression
44	adhalin	AF019564	mouse	U08895	861	1	genes overexpressed in muscle, unknown adipocyte expression
45	amphiphysin 2	U86405	mouse	U87558	142	1	genes overexpressed in muscle, unknown adipocyte expression
46	angiopoietin-related protein	AF153606	human	AF153606	722	1	genes overexpressed in muscle, unknown adipocyte expression
47	cytochrome c oxidase subunit VIII-H precursor (COX8H)	U15541	mouse	none	218	1	genes overexpressed in muscle, unknown adipocyte expression
48	enolase 3, beta	NM_007933	mouse	NM_001976	1171	1	genes overexpressed in muscle, unknown adipocyte expression
49	eukaryotic initiation factor 5 (eIF-5)	L11651	rat	NM_001969	22	1	genes overexpressed in muscle, unknown adipocyte expression
50	GLYCOGEN PHOSPHORYLASE	NM_011224 J03544	mouse	J03544	432	1	genes overexpressed in muscle, unknown adipocyte expression
51	insulin-like growth factor binding protein 5 (IGFBP5)	L12447	mouse	M65062	973	1	genes overexpressed in muscle, unknown adipocyte expression
52	Kif1b; kinesin-like protein	D17577	mouse	AB040881	127	1	genes overexpressed in muscle, unknown adipocyte expression
53	Madr2 (TGFB binding partner)	U60530	mouse	NM_005901	429	1	genes overexpressed in muscle, unknown adipocyte expression
54	methyl-CpG binding protein 2	AF072243	mouse	NM_003927	676	1	genes overexpressed in muscle, unknown adipocyte expression

Figure 7F

55	myosin light chain 2 (check)	U77943	mouse	NM_013292	1029	1	genes overexpressed in muscle, unknown adipocyte expression
56	NADH:ubiquinone oxidoreductase PGIV subunit	AF044953	human	AF044953	632	1	genes overexpressed in muscle, unknown adipocyte expression
57	outer membrane protein (OMP25)	AF107295	rat	NM_018373	937	1	genes overexpressed in muscle, unknown adipocyte expression
58	phosphofructokinase muscle isozyme	U25651	rat	M26066	1068	1	genes overexpressed in muscle, unknown adipocyte expression
59	phospholemman	AF089734	mouse	U72245	865	1	genes overexpressed in muscle, unknown adipocyte expression
60	sarcoglycan, alpha	NM_009161	mouse	NM_000023	842	1	genes overexpressed in muscle, unknown adipocyte expression
61	Skeletal muscle Actin	M12866	mouse	NM_001100	639	1	genes overexpressed in muscle, unknown adipocyte expression
62	skeletal muscle LIM protein (FHL1)	U77039	mouse	U60115	817	1	genes overexpressed in muscle, unknown adipocyte expression
63	tubulin alpha 4 (Tuba4)	NM_009447	mouse	X06956	693	1	genes overexpressed in muscle, unknown adipocyte expression
64	voltage dependent anion channel 2	U30838	mouse	NM_003375	23	1	genes overexpressed in muscle, unknown adipocyte expression
65	Na+ K+-ATPase alpha	M14512	rat	NM_000702	135 440 492 562 570 624 704 746 824 919 1163	11	genes of unknown expression (metabolic)
66	carboxylesterase (Es-HVEL)	X65296	rat	NM_012122	427 494 799 949 1066	5	genes of unknown expression (metabolic)
67	ferrochelatase	NM_007998	mouse	NM_000140	431 578 666 812 1064	5	genes of unknown expression (metabolic)

Figure 7G

68	lactate dehydrogenase 1, A chain (Ldh1)	NM_010699	mouse	NM_005566	496 576 618 687 1172	5	genes of unknown expression (metabolic)
69	long-chain acyl-CoA dehydrogenase	U21489	mouse	M74096	373 628 789 820 1117	5	genes of unknown expression (metabolic)
70	NAD+-isocitrate dehydrogenase alpha subunit	X87172	macaque	U07681	248 587 706 1062 1085	5	genes of unknown expression (metabolic)
71	ATP synthase, H+ transporting, mitochondrial F1F0complex, subunit e (Atp5k)	NM_007507	mouse	NM_007100	559 806 1136 1203	4	genes of unknown expression (metabolic)
72	citrate synthase	AF047042	human	NM_004077	308 490 545	3	genes of unknown expression (metabolic)
73	enoyl-CoA hydratase	X15958	rat	X15958	204 211 891	3	genes of unknown expression (metabolic)
74	NADH-UBIQUINONE OXIDOREDUCTASE Fe-S PROTEIN 2; NDUF2	AF013160	human	AF013160	561 899 1152	3	genes of unknown expression (metabolic)
75	UDP-glucose pyrophosphorylase 2 (UGP2)	NM_006759	human	NM_006759	621 754 800	3	genes of unknown expression (metabolic)
76	2,4-dienoyl-CoA reductase	D00569	rat	NM_001359	501 657	2	genes of unknown expression (metabolic)
77	24-kDa subunit of mitochondrial NADH dehydrogenase	M22756	rat	M22538	1069 1098	2	genes of unknown expression (metabolic)
78	cytochrome c oxidase subunit VIIb (cox7b)	W13383	mouse	NM_001866	77 78	2	genes of unknown expression (metabolic)
79	epoxide hydrolase	Z37107	mouse	NM_001979	120 225	2	genes of unknown expression (metabolic)
80	lactate dehydrogenase 2, B chain (Ldh2)	NM_008492	mouse	NM_002300	515 876	2	genes of unknown expression (metabolic)
81	NADH dehydrogenase (ubiquinone) Fe-S protein 3	NM_004551	human	NM_004551	1010 1157	2	genes of unknown expression (metabolic)

Figure 7H

82	peroxiredoxin I (Prxl)	AF157331	mouse	NM_002574	749 779	2	genes of unknown expression (metabolic)
83	peroxisomal phytanoyl-CoA alpha-hydroxylase (PAHX)	AF023463	mouse	NM_006214	491 677	2	genes of unknown expression (metabolic)
84	phosphoglucomutase	L11694	rat	NM_002633	882 1033	2	genes of unknown expression (metabolic)
85	ubiquinol cytochrome c reductaseRieske (iron-sulfur protein)	M24542	rat	NM_006003	565 1187	2	genes of unknown expression (metabolic)
86	very-long-chain acyl-CoA dehydrogenase	AF017176	mouse	NM_000018	294 489	2	genes of unknown expression (metabolic)
87	3-hydroxyisobutyrate dehydrogenase	J04628	rat	AC007130	736	1	genes of unknown expression (metabolic)
88	3-hydroxyisobutyryl-coenzyme A hydrolase (same as b0019)	U66669	human	NM_014362	1130	1	genes of unknown expression (metabolic)
89	aconitase	af086790	human	AF086790	777	1	genes of unknown expression (metabolic)
90	ATP synthase, H+ transporting, mitochondrial F1 complex, alpha subunit, isoform 1 (Atp5a1)	NM_007505	mouse	NM_004046	1209	1	genes of unknown expression (metabolic)
91	ATP-specific succinyl-CoA synthetase beta	AF058955	human	AF058953	1019	1	genes of unknown expression (metabolic)
92	ATPase, Ca++ transporting, cardiac muscle, fast twitch 1 (Atp2a1)	NM_007504	mouse	U96781	980	1	genes of unknown expression (metabolic)
93	branched chain aminotransferase (BCATm)	U68526	mouse	NM_001190	646	1	genes of unknown expression (metabolic)
94	cytochrome c-1 (CYC1)	NM_001916	human	NM_001916	1147	1	genes of unknown expression (metabolic)

Figure 7I

95	cytosolic class 3 aldehyde dehydrogenase (Adh4)	AF032920	mouse	none	975	1	genes of unknown expression (metabolic)
96	flavin-containing monooxygenase 1	NM_010231	mouse	NM_002021	702	1	genes of unknown expression (metabolic)
97	glycogen synthase	X94616	mouse	NM_002103	787	1	genes of unknown expression (metabolic)
98	Inducible 6-phosphofructo-2-kinase/fructose 2,6-bisphosphatase (IPFK-2) same as b0212	AF056320	human	AF056320	1084	1	genes of unknown expression (metabolic)
99	isopentenyl diphosphate dimethylallyl diphosphate isomerase	AF003836	hamster	NM_004508	723	1	genes of unknown expression (metabolic)
100	malate dehydrogenase	M16229	mouse	NM_005918	12	1	genes of unknown expression (metabolic)
101	malate dehydrogenase, soluble (Mor2).	NM_008618	mouse	NM_005917	590	1	genes of unknown expression (metabolic)
102	methylmalonate semialdehyde dehydrogenase via est	AA497435	mouse	AF159889	932	1	genes of unknown expression (metabolic)
103	mitochondrial long-chain 3-ketoacyl-CoA thiolase	D16479	rat	NM_000183	1072	1	genes of unknown expression (metabolic)
104	NADH dehydrogenase (ubiquinone) 1 alpha	NM_004542	human	NM_004542	541	1	genes of unknown expression (metabolic)
105	NADH dehydrogenase (ubiquinone) 1 beta subcomplex	NM_005004	human	NM_005004	711	1	genes of unknown expression (metabolic)
106	NADH-ubiquinone oxidoreductase B8 subunit (Ndufa2)	AF124786	mouse	NM_002488	782	1	genes of unknown expression (metabolic)
107	NADH-Ubiquinone reductase (Complex I-30KD)	M58469 J05314	cow	NM_004551	187	1	genes of unknown expression (metabolic)
108	NADPH-cytochrome P450 oxidoreductase	D17571	mouse	S90469	731	1	genes of unknown expression (metabolic)

Figure 7J

109	phenol/aryl form sulfotransferase (M-STP1)	L02331	mouse	NM_001055	791	1	genes of unknown expression (metabolic)
110	pyruvate dehydrogenase E1alpha subunit (Pdha1) via est	NM_008810	mouse	NM_000284	972	1	genes of unknown expression (metabolic)
111	SELENOPHOSPHATE SYNTHETASE	NM_012247	human	NM_012247	997	1	genes of unknown expression (metabolic)
112	ubiquinol-cytochrome-c reductase (subunit I)	X59692	cow	D26485	1129	1	genes of unknown expression (metabolic)
113	endophilin b1 isoform (unknown splicing varient)	AA517877 AA462484	mouse	AF263293	139 276 771 1115	4	genes of unknown expression (signaling)
114	Bet 3	AF041433	mouse	NM_014408	59 610 1051	3	genes of unknown expression (signaling)
115	PCTAIRE-motif protein kinase 3	NM_008795	mouse	X66362	743 757 913	3	genes of unknown expression (signaling)
116	Sid470p (CHP calcineurin B homologous protein) or EF-hand Ca(2+)-binding protein p22)	AB025217	mouse	NM_007236	673 862 863	3	genes of unknown expression (signaling)
117	cbl-b	U26710	human	U26710	555 914	2	genes of unknown expression (signaling)
118	Msx-interacting-zinc finger protein 1 (Miz1)	AF039567	mouse	NM_004671	261 1184	2	genes of unknown expression (signaling)
119	protein tyrosine phosphatase PTP alpha/LRP	M36033 M33671	mouse	NM_002836	763 805	2	genes of unknown expression (signaling)
120	quaking type I (QKI) (KH domain RNA binding protein)	U44940 AF091047	mouse	AF142418	742 756	2	genes of unknown expression (signaling)
121	RIP140 (transcriptional activator)	NM_008735	mouse	NM_003489	107 659	2	genes of unknown expression (signaling)
122	sphingomyelin phosphodiesterase	Z14132 Z31654	mouse	X63600	180 1080	2	genes of unknown expression (signaling)

Figure 7K

123	1D-myo-inositol-trisphosphate 3-kinase B	X57206	human	X57206	715	1	genes of unknown expression (signaling)
124	beta-spectrin 2, non-erythrocytic (Spnb2)	NM_009260	mouse	NM_003128	1106	1	genes of unknown expression (signaling)
125	c-abl	K02291	mouse	X51945	767	1	genes of unknown expression (signaling)
126	dishevelled-1 protein (Dvl1)	U28138	mouse	NM_004421	520	1	genes of unknown expression (signaling)
127	inositol polyphosphate-1-phosphatase (Inpp1) (check art for exp)	MMU27295	mouse	NM_002194	1061	1	genes of unknown expression (signaling)
128	Keap1	AB020063	mouse	NM_012289	1026	1	genes of unknown expression (signaling)
129	mamalian homolog of yeast carbon catabolite *repression 4 protein* (transcription factor)	AF183960	mouse	NM_012118	712	1	genes of unknown expression (signaling)
130	polycystic kidney disease-associated protein (PKD1)	L39891	human	L39891	101p	1	genes of unknown expression (signaling)
131	rab28 (check)	X78606	rat	NM_004249	1134	1	genes of unknown expression (signaling)
132	Rab7	NM_009005	mouse	NM_004637	872	1	genes of unknown expression (signaling)
133	rac1 (check abst for exp)	X57277	mouse	M29870	841	1	genes of unknown expression (signaling)
134	Rho-associated coiled-coil forming kinase 1 (Rock1) via est	NM_009071	mouse	NM_005406	1067	1	genes of unknown expression (signaling)
135	SH3-containing protein p4015	AF026505	rat	NM_003603	955	1	genes of unknown expression (signaling)
136	sorting nexin 5 (SNX5)	AF121855	human	AF121855	752	1	genes of unknown expression (signaling)
137	TRIP3 gene; thyroid receptor interactor	L40410	human	L40410	9	1	genes of unknown expression (signaling)

Figure 7L

138	zinc finger protein 216 (Zfp216)	NM_009551	mouse	AF062347	1073	1	genes of unknown expression (signaling)
139	DERP2 and CDLK11H12 protein with weak homology to RAT TEST	AA796646	mouse	AB009685	256 802 1154	3	genes of unknown expression (other)
140	connexin 43 (CX43)	L10388	mouse	M65188	828 939	2	genes of unknown expression (other)
141	deafness dystonia protein (DDP)	U66035	human	U66035	353 364	2	genes of unknown expression (other)
142	mSara	L20294	mouse	AF261717	641 870	2	genes of unknown expression (other)
143	PTP protein	X52101	mouse	X62006	401 468	2	genes of unknown expression (other)
144	putative guanylate binding protein	NM_008620	mouse	NM_002053	568 588	2	genes of unknown expression (other)
145	Rap30 (helicase)	L01267	rat	NM_004128	775 1199	2	genes of unknown expression (other)
146	Aop1 (Mer5) anti-oxidant protein 1	M28723	mouse	NM_006793	226	1	genes of unknown expression (other)
147	aspartyl protease; cathepsin D; protease	X53337	mouse	NM_001909	165	1	genes of unknown expression (other)
148	calpain small subunit (Ca(2+)-dependent proteinase)	NM_009795	mouse	NM_001749	183	1	genes of unknown expression (other)
149	chromatin nonhistone high mobility group protein (HGM-I(Y)	J04179	mouse	NM_002131	585	1	genes of unknown expression (other)
150	complement protein; complement protein C4 and SLP	X05314	mouse	K02404	242	1	genes of unknown expression (other)
151	HC class I H2-L gene Contains Alu sequence	K02896	mouse	none	537	1	genes of unknown expression (other)
152	histone macroH2A1.2 variant (check abst for exp)	AF171080	mouse	AF041483	839	1	genes of unknown expression (other)

Figure 7M

153	JKTBP (hnRNP protein)	AB017020	mouse	D89092	1189	1	genes of unknown expression (other)
154	Mouse heat shock protein 70 cognate	M19141 U73744	mouse	NM_006597	271	1	genes of unknown expression (other)
155	myomegalin (putative PDE-interacting protein)	AF139185	rat	AB00794	727	1	genes of unknown expression (other)
156	neuronatin	NM_010923	mouse	AB002392	557	1	genes of unknown expression (other)
157	procollagen C-endopeptidase enhancer 2	NM_013363	human	NM_013363	1164	1	genes of unknown expression (other)
158	putative RNA helicase and RNA dependent ATPase	AF017153	mouse	NM_001358	446	1	genes of unknown expression (other)
159	ribophorin II (oligosaccharyl transferase (OST) complex protein involved in ER transport)	D31717	mouse	NM_002951	1087	1	genes of unknown expression (other)
160	ribosomal protein L3 homologue	X06323	human	X06323	1054	1	genes of unknown expression (other)
161	sodium-dependent multi-vitamin transporter (SMVT)	AF026554	rat	AF081571	811	1	genes of unknown expression (other)
162	splicing factor, arginine/serine-rich 7 (35kD)	NM_006276	human	NM_006276	795	1	genes of unknown expression (other)
163	sterol carrier protein 2	M62361	mouse	NM_002979	1195	1	genes of unknown expression (other)
164	trans-Golgi network integral membrane protein TGN38.	X53565	rat	NM_006464	1042	1	genes of unknown expression (other)
165	U5 snRNP-specific 200kD protein (RNA helicase)	Z70200	human	Z70200	375	1	genes of unknown expression (other)
166	ubiquitin-conjugating enzyme 7	NM_009456	mouse	NM_003347	614	1	genes of unknown expression (other)

Figure 7N

167	unknown BELONGS TO THE UPF0041 FAMILY (0-44) brain protein 44-like protein (Brp44l)	AA420019 AA590078 AF181116	mouse	AF151887	20 352 526 567 9651131	5	genes of unknown expression (unknown function)
168	G9a (ankyrin-like repeat and set domain possible transcription regulator)	NM_006709	human	NM_006709	208 265 266 813	4	genes of unknown expression (unknown function)
169	bromodomain-containing protein BP75, homologous to CEC01H6 (has a zinc finger and is homologous to transcription factors)	W44243	mouse	AF179285	384 385 573	3	genes of unknown expression (unknown function)
170	thioredoxin interacting factor	AF173681	mouse	NM_006472	790 856	2	genes of unknown expression (unknown function)
171	unknown human protein with homology to mosin and chromosomal binding proteins. Protein has a leucine zipper	AB011166	human	AB011166	101v 110	2	genes of unknown expression (unknown function)
172	unknown protein RS21-C6	AF110764	mouse	none	25 85	2	genes of unknown expression (unknown function)
173	breakpoint cluster region protein 1 (Bcrp1-pending)	NM_011793	mouse	AF044773	550	1	genes of unknown expression (unknown function)
174	H-rev107	X76453	rat	NM_007069	658	1	genes of unknown expression (unknown function)
175	Human alpha3a intergrin binding protein WD40 repeats 5' end homology*	AB023234	human	AB023234	283	1	genes of unknown expression (unknown function)

Figure 7O

176	insulin induced protein 2	AF125392	human	AF125392	1139	1	genes of unknown expression (unknown function)
177	KIAA0304 gene (homologus to trithorax and HRX genes)	AB002302	human	AB002302	416	1	genes of unknown expression (unknown function)
178	low density lipoprotein receptor related protein associated protein	S67967	mouse	NM_002337	83	1	genes of unknown expression (unknown function)
179	mab-21	U38810	human	U38810	449	1	genes of unknown expression (unknown function)
180	Mus musculus cDNA sequence AF18047 unknown gene	NM_015763	mouse	D80010	529	1	genes of unknown expression (unknown function)
181	PEROXISOMAL FARNESYLATED PROTEIN (PxF)	U05959	hamster	NM_002857	1104	1	genes of unknown expression (unknown function)
182	Pyruvate dehydrogenase complex, lipoyl-containing component X; E3-binding protein (PDX1)	NM_003477	human	NM_003477	518	1	genes of unknown expression (unknown function)
183	RNaseP protein p38 (RPP38)	U77664	human	U77664	1132	1	genes of unknown expression (unknown function)
184	TEB4 protein	AF009301	human	AF009301	952	1	genes of unknown expression (unknown function)
185	unknown gene homologous to Homo sapiens clone H41 unknown mRNA.	AA407977 AF103803	human	AF103803	784	1	genes of unknown expression (unknown function)
186	unknown homologous to Homo sapiens CGI-51 protein mRNA	AF151809	human	AF151809	1202	1	genes of unknown expression (unknown function)

Figure 7P

187	unknown homologous to HYPOTHETICAL PROTEIN KIAA0193	AI006170 NM_014766	human	NM_014766	990	1	genes of unknown expression (unknown function)
188	unknown homology to Homo sapiens antigen NY-CO-10 (#AF039692) has PEPTIDYL-PROLYL CIS-TRANS ISOMERASE domain possible cyclophilin	AW120624	human	AW120624	414	1	genes of unknown expression (unknown function)
189	unknown human gene FLJ10530 homology to Sulfatase Metallo-beta-lactamase superfamily	Ak001392	human	Ak001392	678	1	genes of unknown expression (unknown function)
190	unknown human gene FLJ20420 fis with coil coil domain	AK000427	human	AK000427	675	1	genes of unknown expression (unknown function)
191	unknown	AA286376 AA79654			89 105 579 603 664 680 1160	7	unknown genes with est
192	unknown	AA240742			729 958 1077	3	unknown genes with est
193	unknown (maybe 112)	AA389142			121 203 864	3	unknown genes with est
194	similar to type II membrane protein (tazarotene-induced gene 2 (TIG2))	AB015632			591 1089	2	unknown genes with est
195	unknown	AA572287			124 549	2	unknown genes with est
196	unknown	AA422249			222 979	2	unknown genes with est
197	unknown	AA105146			380 390	2	unknown genes with est
198	unknown	AA880575			443 460	2	unknown genes with est
199	unknown	AA53681			442 1044	2	unknown genes with est
200	Unknown Homologous To Clone 584089 5' Similar To Tr:G550452 G550452 3 Methylcrotonyl-Coa Carboxylase Biotin-Carrier Domain	AA125589			103 1108	2	unknown genes with est

Figure 7Q

201	unknown homologous to KIAA0907 unknown human protein with KH domain (RNA binding domain)	AI117411		974 978	2	unknown genes with est
202	unknown homology to kinesin like protein from C elegans	AA065565		859 986	2	unknown genes with est
203	unknown in contiq in Chi du Chat transcription region	AA457865		250 535	2	unknown genes with est
204	unknown maybe SIN3A (NM_011378)	AI746380		762 854	2	unknown genes with est
205	unknown tc homology to crgA (ring/zind finger protein) (no mouse est only human ests)	AA402287		596 640	2	unknown genes with est
206	similar to drosphophila putative type III alcohol dehydrogenase (T3dh)	U66411		1217	1	unknown genes with est
207	uknown possibley homologous to unkown human protein similar to myotubularin dual specificity phosphatase with a FYVE domain	AV154597 AA027890		793	1	unknown genes with est
208	unknown	AA472125		196	1	unknown genes with est
209	unknown	AI551808		600	1	unknown genes with est
210	unknown	AA254592		504	1	unknown genes with est
211	unknown	BB195833		689	1	unknown genes with est
212	unknown	AI849511		684	1	unknown genes with est
213	unknown	AI552298		681	1	unknown genes with est
214	unknown	AI836168		748	1	unknown genes with est
215	unknown	AA230655		846	1	unknown genes with est

Figure 7R

216	unknown	AA422762		901	1	unknown genes with est
217	unknown	AI194220		884	1	unknown genes with est
218	unknown	AI613609		1032	1	unknown genes with est
219	unknown	AW210669		996	1	unknown genes with est
220	unknown	AI118610		935	1	unknown genes with est
221	unknown	AA798041		1099	1	unknown genes with est
222	unknown	AW744318		1144	1	unknown genes with est
223	unknown	AA880976		1148	1	unknown genes with est
224	unknown	AA472346		1176	1	unknown genes with est
225	unknown	AA823731		1201	1	unknown genes with est
226	unknown	AI158578		1205	1	unknown genes with est
227	unknown (3' end)	AI853437		785	1	unknown genes with est
228	unknown (check for other end)	AI042737		991	1	unknown genes with est
229	unknown est 5' similar to WP:T19C3.4 CE02055 POLYPOSIS LOCUS PROTEIN 1 potential transmembrane protein	AA797245		39	1	unknown genes with est
230	unknown (genomic sequence from mouse 9)	AA472177		538	1	unknown genes with est
231	unknown (homologous to human unknown human gene HSPC154)	AA986310		536	1	unknown genes with est
232	unknown homologous to unknown human clone 24640	AW209647		724	1	unknown genes with est
233	unknown (homologous to unknown human gene MAC30)	AA003926		513	1	unknown genes with est

Figure 7S

234	unknown (homologus to unknown human clone CGI 126)	AI526927		533	1	unknown genes with est
235	unknown homology to an exon of sel-1 via est protein with FN2 domain (Homo sapiens gene for TSA305)	AV117710		983	1	unknown genes with est
236	unknown homology to ANON-66Db	AA190077		439	1	unknown genes with est
237	unknown (homology to CG9921 gene product or P100 protease)	AA518356		1071	1	unknown genes with est
238	unknown (homology to DKFZp586K1424)	AA210389		564	1	unknown genes with est
239	unknown homology to *unknown human protein* KIAA1098 protein with *ring finger, b-box and* SPRY domains (possible *DNAJ domain) homology* to Frog transcription factor xnf7	AA500111		943	1	unknown genes with est
240	unknown homology via est vr14c09 to kineasin light chain	AA498553		76	1	unknown genes with est
241	unknown (maybe a repeated element	AA920762		692	1	unknown genes with est
242	unknown (maybe importin)	AA422297		1112	1	unknown genes with est
243	unknown PHD finger protein homologous to human gene KIAA0783	AA414352		236	1	unknown genes with est
244	unknown possible homology to myosin	AA517877		152	1	unknown genes with est

Figure 7T

245	unknown possible phospholipase A2 protein	AA543736		635	1	unknown genes with est
246	unknown possible repetative element	AA120704		794	1	unknown genes with est
247	unknown possible repetive element	AA273814		823	1	unknown genes with est
248	unknown possibly homolgous to drosophila gene CG12848	W13360		1023	1	unknown genes with est
249	unknown possibly spermatogenesis associated PD1 (bBox possible transcription factor)	AA571641		580	1	unknown genes with est
250	unknown (potential zinc finger protein) with homolgoy to autocrine motilityfactor receptor????	AW107472		507	1	unknown genes with est
251	unknown related to mbln protein and mindblind	AA511443		367	1	unknown genes with est
252	unknown similar ot mitochondrial carrier proteins	AA474460		739	1	unknown genes with est
253	unknown (similar to human to KIAA0112)	AA210389		572	1	unknown genes with est
254	unknown similar to SW:YKT5_YEAST P36046 HYPOTHETICAL 47.4 KD PROTEIN	AA616045		95	1	unknown genes with est
255	unknown similar to unknown human cDNA DKFZp761D0223similar to rb gaps	AI503800		582	1	unknown genes with est

Figure 7U

256	unknown similar to unknown human protein KIAA0664 (Clu1 like)	AV043819			840	1	unknown genes with est
257	unkown gene homologous to unknown human gene HSPC241	AA500671			710	1	unknown genes with est
258	unknown				827 984	2	unknown genes without est
259	unknown				148	1	unknown genes without est
260	unknown				1083	1	unknown genes without est
261	mitochondrial genome	V00711	mouse	NC_001807	47 132 155 160 190 199 220 230 234 235a 240 275 289 312 343 356 393 417 423 425 434 498 502 505 517 524 539 540 546 574 575 586 599 611 612 617 622 627 647 648 651 661 665 674 679 696 679 699 700 705 714 725 726 728 737 738 741 760 761 783 788 801 804 810	107	mitochondrial
262	Gas????	AF116268	mouse	X04409	552 682	2	genes not overexpressed
263	Bach1 via TC81197 (transcription factor)	NM_007520	mouse	NM_001186	544	1	genes not overexpressed
264	retinoblastoma-binding protein (mRbAp48)	U35141	mouse	NM_006510	1017	1	genes not overexpressed
265	skeletal muscle alpha tropomyosin	X64831	mouse	NM_000366	903	1	genes not overexpressed

Figure 8A

Primary Adipocytes - PEAK 1

Band	Protein/Accession number
1	AM-2 Receptor (49942), human (NP_002323)
2	AM-2 Receptor (49942), human (NP_002323)
3	AM-2 Receptor (49942), human (NP_002323)
4	Non-Identified
5	IGFII-Receptor/ Cation-independent Mannose 6-P Receptor (1398935), human (NP_000867)
6	Fatty Acid Synthase (66561), human (G01880); Acetyl-CoA Carboxylase (116670), human (NP_000655)
7	DNAJ-Domain containing protein (3327170), human (T00361)
8	ABC1-Protein (6005701), human (009099)
9	IRAP (1674503), human (CAB61646)
10	IRAP (1674503), human (CAB61646)
11	IRAP (1674503), human (CAB61646)
12	Non-Identified
13	ATP Citrate-Lyase (113116), human (AAB60340)
14	Amine Oxidase (4185817), human (NP_0037250)
15	Amine Oxidase (4185817), human (NP_0037250)
16	Alpha-2 Macroglobulin Receptor (4758686), human (NP_002323); Hormone sensitive Lipase (1346458), human (Q05469)
17	CD36 (3273897), human (NP_000063)
18	CD36 (3273897), human (NP_000063)
19	Long Chain Acyl-CoA Synthetase (126011), human (P33121)

Figure 8B

20 Long Chain Acyl-CoA Synthetase (126011), human (P33121)

21 Phosphatidylserine-Binding Protein/Substrate of PKC (455719), human (NP_004648)

22 EH-Domain-Containing Protein (1861774), human (1861774)

23 Novel

24 Carboxylesterase (1407780), human (AAB03611);
Vimentin (401365), human (A25074)

25 α-Tubulin (223556), human (A23035);
GLUT4 (121763), human (NP_001033);
DLAST (266684), human (AAD30181)

26 GLUT4 (121763), human (NP_001033)

27 GLUT4-P (121763), human (NP_001033)

28 GLUT4 (121763), human (NP_001033);
Non-Identified

29 Junctional Adhesion Molecule (5457119), human (AAD3794)

30 SCAMP (5032077), human (NP_005689)

31 Ribosomal Protein L6 (2507315), human (Q02878)

32 SCAMP (5032077), human (NP005689)

33 Non-Identified

34 Non-Identified

35 29 kD-Golgi SNARE (3213227), human (NP_006361);
Non-Identified

36 Non-Identified

37 Caveolin-1 (1705645), human (AAD23745);
Non-Identified

38 Caveolin-1 (1705645), human (AAD23745);

Figure 8C

39 Novel

40 Novel

3T3-L1 Adipocytes - PEAK 1

Band	Protein/Accession number
1	AM-2 Receptor (4758686), human (CAA38905) Myeloid associeted differentiation protein (3212400); human (3212400)
2	IGF II- Receptor (1709091), human (NP_000867) Non-muscle myosin heavy chain A (967249), human (P35579)
3	α-1 subunity Na+/K+-ATPase (114374), human (P05023) SERCA1a Ca2+-ATPase (477339), human (1586563)
4	IRAP (2144020), human (CAB61646) Leucine aminopeptidase placental (1888354), human (CAB94753) Protein 4.1G (3064263), human (P11171) Glutamyl-prolyl-t-RNA (4758294), human (NP_004437) ES/130 (4759056), human (NP_004578)
5	Coatomer protein α-subunit (6642754), human (NP_004362)
6	Coatomer protein α-subunit (6642754), human (NP_004362)
7	Non-Identified
8	Pyruvate Carboxylase (6679237), human (NP_000911)
9	Sortilin 1 (6653197), human (NP_002950) Myosin I heavy chain (480659), human (CAA67131) Proteasome subunit p112 (4506225), human (NP_002798) Major vault protein (497940), human (NP_005106)
10	Sortilin 1 (4507159), human (NP_002950) α-Catenin (2134736), human (I39438) Protein C23 (128843), human (NP_005372) Major vault protein (497940), human (NP_005106)

Figure 8D

11	Coatomer protein β2-subunit (4758032), human (NP_004757) AP2 β1-subunit (4557469), human (NP_001273) AP2 α2-subunit (6671563), human (AAD15564) LIMP-II (126381), human (NP_002285) GRP94 (119362), human (AAF82792) Alix (3550456), human (NP_037506) Proteasome subunit p97 (1060888), human (BAA11226)
12	Amine oxidase (5902787), human (NP_003725) Long chain Acyl-CoA synthetase 2 (6679739), human (P33121) Geosolin (121117), human (NP_000168) Calnexin (6671664), human (P27824) STA1(6678153), human (NP_009330) Hsp 90 (6680307), human (NP_031381) Glycogen synthase (517112), human (BAA06154)
13	RET II (3551509), human (NP_005104) Hormone sensitive lipase (1708847), human (Q05469)
14	Moesin (4505257), human (NP_002435) Long chain Acyl-CoA synthetase 2 (6679739), human (P33121) Calcium binding protein 2 (729436), human (NP_004902) Cytochrome P450 (6679421), human (P16435) GRP78 (121574), human (CAB71335)
15	Protein 4.1B (5020274), human (NP_001422) S3-12 protein (3236368), human (NP_001113) β1-Integrin (124964), human (P05556)
16	α6-Integrin (3183038), human (P23229) Vitronectin receptor (6680486), human (NP_002201) Oncoprotein LFC (6678666), human (Q92974) eIF3-p110 (4503525), human (AAC27674)
17	Calcium binding protein 63K (2493471), human (NP_006175) Dihydrolipoamide acetyltransferase (226207), human (226207) Primary biliary cirrhosis autoantigen (2117706), human (P10515) Complement C3 (1352102), human (NP_000055)

Figure 8E

18 Vimentin (281012), human (A25074)
Carboxylesterase 1 (6679689), human (NP_036254)
Lipoprotein lipase (6678710), human (NP_000228)
PTRF (6679567), human (AAC63404)
Prolyl-4-hidroxylase β-subunit (2507460), human (P07237)
ER-60 protease (927670), human (P30101)

19 Dynein light intermediate chain 53/55 (2618478), human (NP_006132)
Calcium binding protein 1 (2501206), human (NP_005733)
Dihydrolipoamide acetyltransferase (1710279), human (AAB50223)
DEBT-91 (4838557), human (AAF67009)
α-subunit ATP-synthase (6680748), human (NP_004037)

20 GLUT4 (6678015), human (NP_001033)
Actin α-2(4501883), human (NP_001604)
Ribosomal protein L3 (4506649), human (NP_000958)

21 SCAMP (3914963), human (O14828)
GLUT4 (6678015), human (NP_001033)

22 Annexin V(1351942), human (999924)
Glyceraldehyde 3-phosphate dehydrogenase(4176768), human (NP_002037)
TAX (1350763), human (Q02878)
Ribosomal protein L5 (1173056), human (NP_000960)

23 Proteasome C2 subunit (5757653), human (NP_002777)
Proteasome 26S subunit PP31 (4506233), human (NP_002803)
Ribosomal protein L7A (4506661), human (NP_000963)
Ribosomal protein S3 (200770), human (AAA03081)
Heme oxigenase 1 (123447), human (NP_002124)

24 BAP-31 protein (2137162), human (NP_005736)
14-3-3 protein (4507953), human (NP_003397)
Prohibitin (4505773), human (NP_002625)
Micropain subunit IOTA (296736), human (CAA43964)
Proteasome subunit zeta chain (4506187), human (NP_002781)
C6-1 Proteasome PSMA7 chain (3805978), human (NP_002783)

Figure 8F

25 v-SNARE Vti1-b (3213229), human (NP_006361)
SNAP-23/Syndet (6678049), human (O00161)
BAP-31 protein (213712), human (NP_005736)
Ubiquitin precursor (2118964), human (I52220)
Proteasome subunit zeta chain (4506187), human (NP_002781)
C6-1 Proteasome PSMA7 chain (3805978), human (NP_002783)
Ribosomal protein L13 (1083788), human (NP_000968)

26 Synaptobrevin-like protein (1617398), human (NP_005629)
Ferritin light chain (120524), human (P02792)
Ribosomal protein (6677781), human (NP_000983)

27 BET-3 (2791806), human (NP_055223)
Myosin regulatory light chain (NP_059039), human (AAA67367)
Type II membrane protein (), human ()
Polyposis locus protein 1 homolog (Q60870), human (Q00765)
Cytochrome B5 (117809), human (P00167)

28 ATP-Synthase, δ-subunit (4502297), human (NP_001678)
Cytochrome B5 (117809), human (P00167)
Mytochondrial import receptor subunit TOM20 (298698), human (NP_055580)
Ribosomal protein 40S (4506695), human (NP_001013)

29 VAMP3 (6678553), human (NP_004772)
Cystatin C (1345935), human (AAA52164)
Histone H1b (356168), human (NP_005312)

30 Non-muscle Myosin light chain 6 (127148), human (P24572)
Cytochrome B5 (231928), human (P00167)
Ribosomal protein 40S (4506695), human (NP_001013)

31 Annexin V(6753060), human (999924)
Lipocortin V(2981437), human (999924)
Membrane glicoprotein GP42 (114835), human (NP_001719)
Ribosomal protein S6 (225901), human (P10660)

32 Glyceraldehyde 3-phosphate dehydrogenase(120707), human (NP_002037)
Calpactin I (113951), human (NP_004030)
Ribosomal protein L5 (1173056), human (NP_000960)

33 β-Tubulin (135451), human (T08726)
IMPDH II (124427), human (P12268)
Vacuolar H+-ATPase (522193), human (AAA58661)

Figure 8G

34 α-1 subunity Na+/K+-ATPase (358959), human (P05023)
Proteasome 26S p97 (1060888), human (BAA11226)
Alix (3550456), human (NP_037506)
Gelsolin (121117), human (NP_000168)
β-Catenin (4503131), human (NP_001895)

35 AM-2 precursor (4557225), human (NP_000005)
Ca2+-ATPase 1-plasma membrane (4502287), human (NP_001673)
α-1 subunity Na+/K+-ATPase (358959), human (P05023)

3T3-L1 Adipocytes - PEAK 2

Band Protein/Accession number

1 Dynein heavy chain (294543), human (Q14204)
Plectin (1709655), human (CAA91196)

2 MAP4 (7106363), human (NP_002366)
HsGCN1 (3970973), human (AAC83183)
Telomerase protein 1 (6678285), human (NP_009041)

3 Myosin heavy chain A (6981236), human (P35579)
MAP4 (7106363), human (NP_002366)

4 IQ-motif containing GTPase activatin protein (4506787),human (NP_003861)
MAP4 (7106363), human (NP_002366)

5 Ribosomal binding protein ES/130 (4759056), human (NP_004578)

6 Kinesin-related protein (2370435), human (NP_006603)
Isoleucine t-RNA synthetase (4504555), human (NP_002152)

7 Pyruvate carboxylase (6679237), human (NP_000911)
Serine/arginine-rich protein specific kinase (6678135),human (AAC29140)
Ribonucleoprotein U (4758546), human (NP_004492)

8 Glycogen synthase (6680141), human (P13807)

Figure 8H

9	Hsp70 protein (5729877), human (NP_006588) Long chain fatty acid Acyl-CoA synthetase (126011),human (P33121)
10	Hsp70 protein (497940), human (NP_006588) GRP78 (121567), human (P11021) Estrogen-responsive finger protein (2137285), human (NP_005073) Major vault protein (497940), human (NP_005106)
11	Vimentin (138536), human (A25074) PI-Phospholipase I (91897), human (JC5704) DJ149A16.6 (Novel Protein) (4468866), human (HSPC117)
12	26S Protease regulatory subunit 6A (2492523), human (P17980) Human elongation factor 1-alpha (4503471), human (NP_001393) Ribosomal protein L4 (1710511), human (BAA04887)
13	p38-2G4, Cell cycle protein (1083448), human (AAD05561) Ribosomal protein L3 (7305441), human (NP_000958) TAX responsive enhancer (6755354), human (Q02878)
14	Nucleolar phosphoprotein (7242160), human (P06748) Endothelial monocyte-activating protein II (6679639), human (B55053)
15	Pyruvate dehydrogenase (2144337), human (DEHUPB) Ribosomal protein L6 (2507315), human (Q02878) Ribosomal protein L5 (1173056), human (NP_000960)
16	Ribosomal protein L6 (6755354), human (Q02878) Ribosomal protein S4 (227229), human (NP_000998)
17	Ribosomal protein 1 10A (6755350), human (CAB38627) Ribosomal protein S8 (4506743), human (NP_001003) BAP-31 (2137162), human (S49265) Ribosomal protein L13 (1350662), human (NP_000968)
18	Ribosomal protein S14 (133785), human (NP_005608) Ribosomal protein S19 (4506695), human (NP_001013) Ribosomal protein S26 (6981488), human (P02383) Ribosomal protein L31 (4506633), human (NP_000984) Ribosomal protein S25 (4506707), human (NP_001019) Cytochrome B5 (231928), human (P00167) H2B histone family member (4504275), human (NP_003518)

Figure 8I

19 Histone H 1.4 (1170151), human (NP_005312)
Ribosomal protein S2 (4506719), human (P15880)

20 Aspartate t-RNA ligase (135099),human (NP_001340)
Dynein light intermediate chain 2 (2494218), human (NP_006132)
26S Protease regulatory subunit 6A (2492523), human (P17980)
Tubulin β-5 (7106439),human (NP_006078)

Figure 9

Protein/Gene	Library	Primary Adipocyte	3T3-L1 Peak 1	3T3-L1 Peak 2
Glut4	Union	Yes	Yes	Yes
Long-chain Acyl-CoA dehydrogenase(synthetase)	Both	Yes	Yes	Yes
Acetyl-coa Carboxylase	Subtractive	Yes	?	?
Carboxylesterase	Union	Yes	?	?
Caveolin-1	Both	Yes	?	?
CDC36	Both	Yes	?	?
Glyceraldehyde-3-Phosphate Dehydrogenase	Union	?	Yes	Yes
Glycogen Synthase	Union	?	Yes	Yes
Grp78	Subtractive	?	Yes	Yes
Grp94	Subtractive	?	Yes	Yes
Pyruvate Carboxlase	Subtractive	?	Yes	Yes
Ribosomal Protein L3	Union	?	Yes	Yes
Bet3	Union	?	Yes	?
Lipoprotein Lipase	Union	?	Yes	?
Myosin Regulatory Light Chain	Union	?	Yes	?
Na+ K+-ATPase alpha	Union	?	Yes	?
S3-12	Union	?	Yes	?
Heat Shock Protein 70	Union	?	?	Yes

Figure 10

c0582

Fig. 10A Fauchere/Pliska hydrophobicity: Window = 11

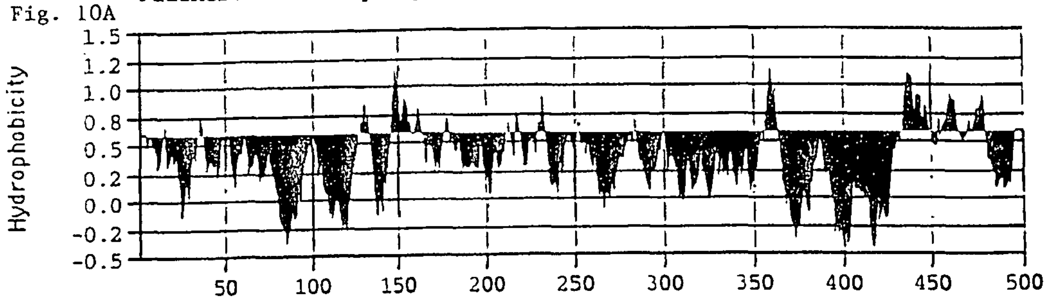

Fig. 10B Janin hydrophobicity: Window = 11

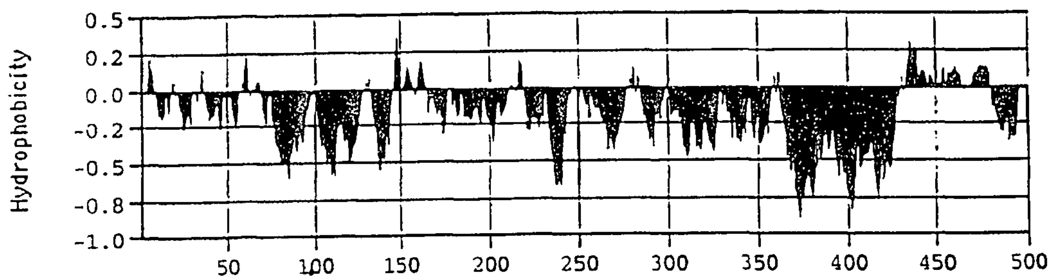

Fig. 10C Manavalan hydrophobicity: Window = 11

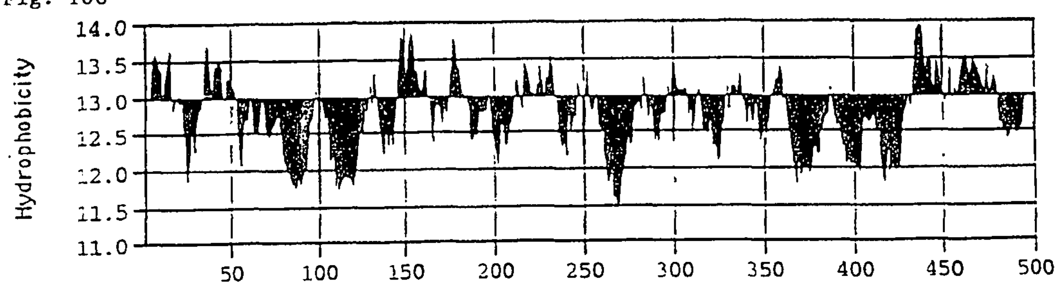

Fig. 10D Sweet/Eisenberg hydrophobicity: Window = 11

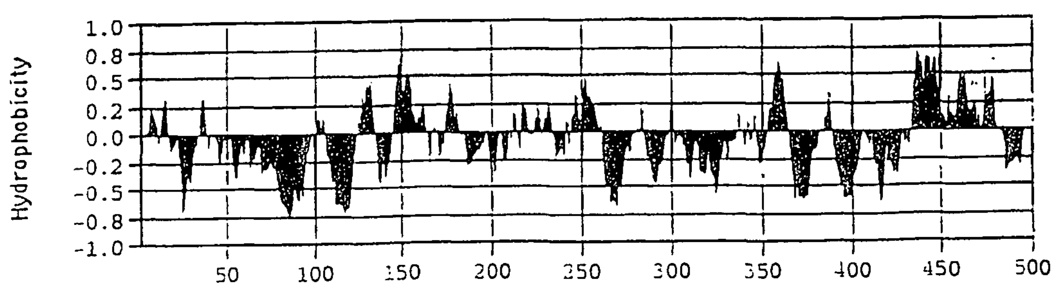

Figure 11

c0139

Fig.11A Faucnere/Pliska hydrophobicity: Window = 11

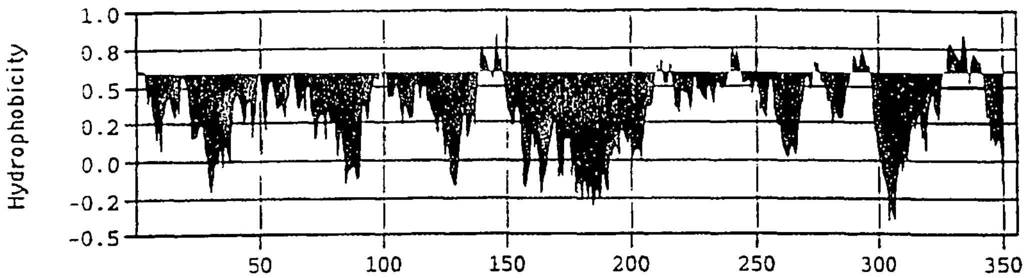

Fig. 11B Janin hydrophobicity: Window = 11

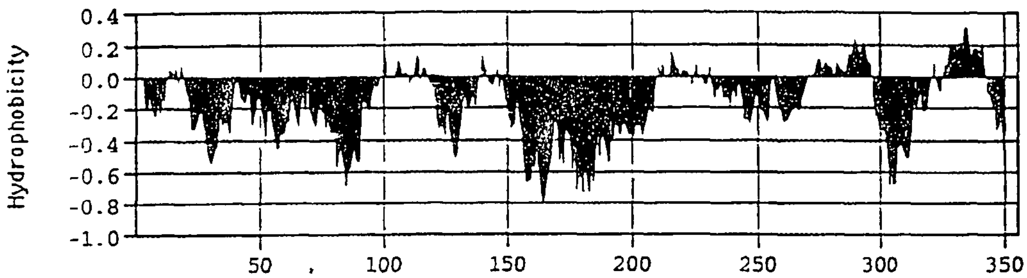

Fig. 11C Manavalan hydrophobicity: Window = 11

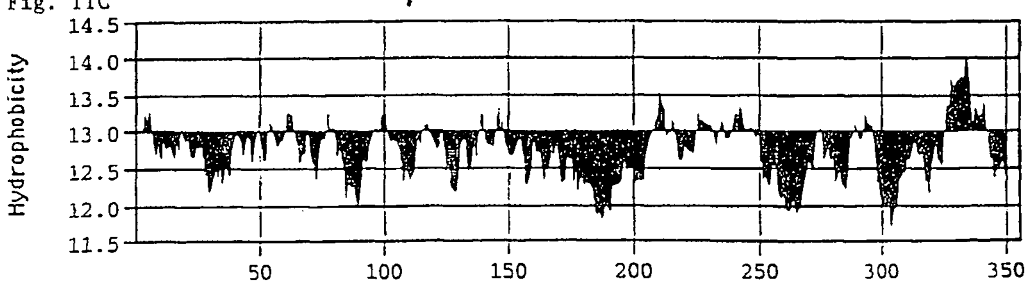

Fig. 11D Sweet/Eisenberg hydrophobicity: Window = 11

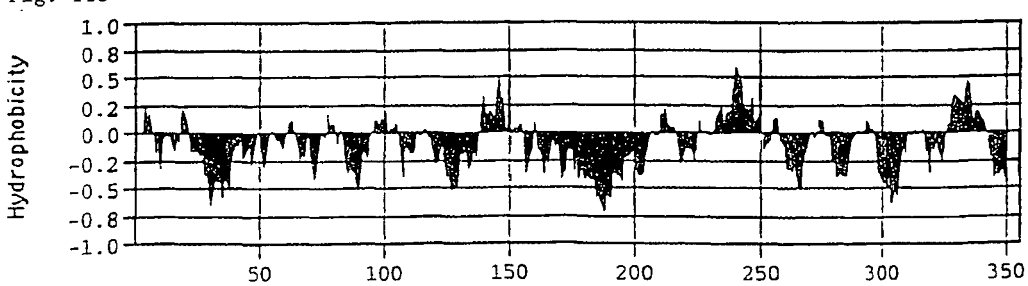

Figure 12

b0175

Fig. 12A Fauchere/Pliska hydrophobicity: Window = 11

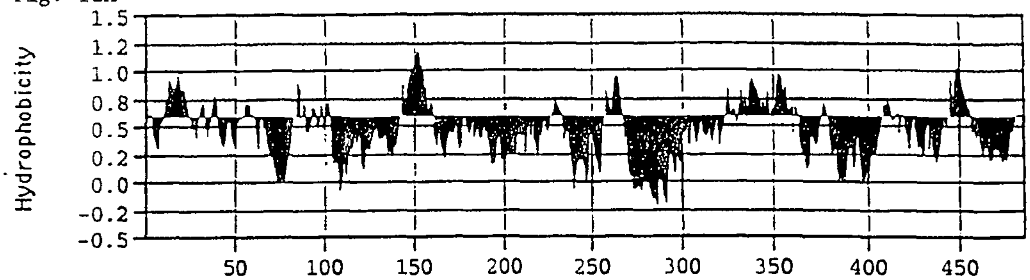

Fig. 12B Janin hydrophobicity: Window = 11

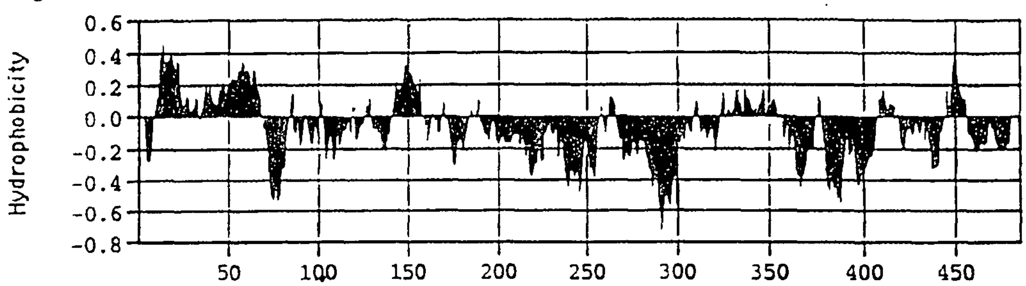

Fig. 12C Manavalan hydrophobicity: Window = 11

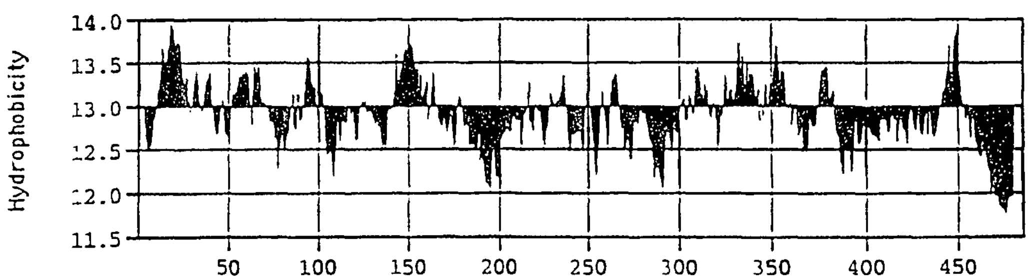

Fig. 12D Sweet/Eisenberg hydrophobicity: Window = 11

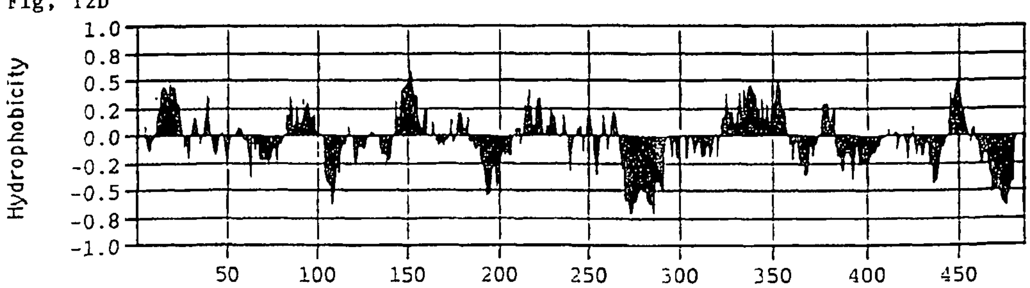

Figure 13A

	Array Name	Probe Set	gene	Accession number	f1	f2	f3	a1	a2	a3	m1	m2	m3	class
1	*MG-U74A*	*98966_at*	*nuclear-encoded mitochondrial acyltransferase*	*L42996*	A	A	A	P	P	P	P	P	P	1
2	*MG-U74B*	*115824_at*	*2,4-dienoyl CoA reductase 1*	*L26050*	A	A	A	P	P	P	P	P	P	1
3	MG-U74A	92592_at	Glycerolphosphate dehydrogenase 1, cytoplasmic	M25558	A	A	A	P	P	P	P	P	P	1
4	MG-U74A	92642_at	carbonic anhydrase II (CAII)	M25944	A	A	A	P	P	P	P	P	P	1
5	MG-U74A	94275_at	Uroporphyrinogen decarboxylase	U34691	A	A	M	P	P	P	P	P	P	1
6	MG-U74A	95645_at	ISCU2	AI837440	A	A	A	P	P	P	P	P	P	1
7	MG-U74A	96095_i_at	Estradiol 17 Beta-Dehydrogenase 4	AI648018	A	A	A	P	P	P	P	P	P	1
8	MG-U74A	101991_at	flavin-containing monooxygenase	D16215	M	A	A	P	P	P	P	P	P	1
9	MG-U74A	102314_at	Glut4	M23383	A	A	A	P	P	P	P	P	P	1
10	MG-U74A	103297_at	6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase	X98848	A	A	A	P	P	P	P	P	P	1
11	MG-U74B	107117_at	peptide methionine sulfoxide reductase	AJ242973	A	A	M	P	P	P	P	P	P	1
12	MG-U74B	110280_at	peroxisomal trans 2-enoyl CoA reductase	AF232011	A	A	A	P	P	P	P	P	P	1
13	MG-U74B	112871_at	5-oxo-L-prolinase	U70825	A	A	A	P	P	P	P	P	P	1
14	MG-U74C	132325_f_at	1,4-alpha-glucan branching enzyme	AAA58642	A	A	A	P	P	P	P	P	P	1
15	MG-U74C	132987_f_at	malate dehydrogenase	AAA37423	A	A	A	P	P	P	P	P	P	1
16	MG-U74A	92202_g_at	ZINC FINGER PROTEIN PLZ	AI553024	A	A	A	P	P	P	P	P	P	2
17	MG-U74A	94365_at	Histidine Triad Nucleotide-Binding Protein	AA710175	A	A	A	P	P	P	P	P	P	2
18	MG-U74A	102366_at	Resisitn	AA718169	A	A	A	P	P	P	P	P	P	2
19	MG-U74A	102958_at	EBF-2	U82441	A	A	M	P	P	P	P	P	P	2
20	MG-U74A	104098_at	peroxisome membrane protein	L28835	A	A	A	P	P	P	P	P	P	2
21	MG-U74B	105496_at	heat shock transcription factor 2	XM_004466	A	A	A	P	P	P	P	P	P	2
22	MG-U74B	108283_at	calcium-independent phospholipase A2	AB041261	A	M	A	P	P	P	P	P	P	2
23	MG-U74B	111708_at	Lpin1	AF180471	A	A	A	P	P	P	P	P	P	2
24	MG-U74B	111925_at	semaphoring	AB029496	A	A	A	P	P	P	P	P	P	2
25	MG-U74B	113344_at	CAP	NM_015385	A	A	A	P	P	P	P	P	P	2
26	MG-U74C	130718_at	tissue inhibitor of metalloproteinases	U76456	M	M	M	P	P	P	P	P	P	2
27	*MG-U74B*	*109963_f_at*	*mitochondrial uncoupling protein 1*	*BC001075*	A	A	A	P	P	P	P	P	P	3
28	MG-U74B	108289_at	keratin 6	L42583	A	A	A	P	P	P	P	P	P	3
29	MG-U74B	109254_f_at	AP-2 50 kDA PROTEIN	U27106	A	A	A	P	P	P	P	P	M	3
30	MG-U74B	111431_at	alpha-actinin 1	M95178	A	A	A	P	P	P	P	P	P	3
31	MG-U74A	95453_f_at	calcium-binding protein S100A1	AF087687	A	A	A	P	P	P	P	P	P	4
32	MG-U74A	99671_at	adipsin	X04673	A	A	A	P	P	P	P	P	P	4

Figure 13B

33	MG-U74A	101575_i_at	Serine protease inhibitor 1	M25529	A	M	A	P	P	P	P	P	P	4
34	MG-U74B	106600_at	PINCH protein	U09284	A	A	A	P	P	P	P	P	P	4
35	MG-U74B	109884_f_at	CD36	L23108	A	A	A	P	P	P	P	P	P	4
36	MG-U74A	94012_at	UI-M-AH1-agv-f-06-0-UI.s1 Mus musculus cDNA	AI848685	A	M	M	M	P	P	P	P	P	5
37	MG-U74A	95596_at	unknown	AW122957	A	A	A	P	P	P	P	P	P	5
38	MG-U74A	96122_at	TCP-1 epsilon	AW049373	A	A	A	P	P	P	P	P	P	5
39	MG-U74A	96237_at	unknown	AI118905	A	A	A	P	P	P	P	P	P	5
40	MG-U74A	97397_at	unknown	AI848344	A	A	A	P	P	P	P	P	P	5
41	MG-U74A	97511_at	unknown	AI846600	A	A	A	P	P	P	P	P	P	5
42	MG-U74A	100458_at	clone E161	X61450	M	A	A	P	P	P	P	P	P	5
43	MG-U74A	104567_at	C15orf4	AI848526	A	A	A	P	P	P	P	P	P	5
44	MG-U74B	105993_at	PTD015 protein	BC00272	A	A	A	P	P	P	P	P	P	5
45	MG-U74B	106215_at	unknown	AK018608	A	A	A	P	P	P	P	P	P	5
46	MG-U74B	106606_at	hypertension-related protein	U41803	A	A	A	P	M	M	P	P	P	5
47	MG-U74B	108493_at	ORF1	AJ278735	A	A	A	P	P	P	P	P	P	5
48	MG-U74B	108499_at	CG13865	AAF45339	A	A	M	P	P	P	P	P	P	5
49	MG-U74B	109124_at	KIAA1288	AB033114	A	A	A	P	P	P	P	P	P	5
50	MG-U74B	110826_at	HSPC202	AF151036	A	A	A	P	P	P	P	P	P	5
51	MG-U74B	112215_at	unknown	AK007799	A	A	A	M	M	M	P	P	P	5
52	MG-U74B	113079_at	FLJ13194	AK023256	A	A	A	P	P	P	P	P	P	5
53	MG-U74B	113599_at	BACR7C10	CAB58018	A	A	A	P	P	P	P	P	P	5
54	MG-U74B	113902_at	HSPC202	AF151036	A	A	A	P	P	P	P	P	P	5
55	MG-U74B	114296_at	FLJ10378	AK001240	A	A	A	P	P	P	P	P	P	5
56	MG-U74B	114491_f_at	unknown	AK014834	A	A	M	P	P	P	P	P	P	5
57	MG-U74B	114801_at	FLJ21893	AK025546	A	A	A	P	P	P	P	P	P	5
58	MG-U74B	115341_at	hypoxia induced gene 1	AF141312	A	A	A	P	P	P	P	P	P	5
59	MG-U74B	115480_at	unknown	AK002741	A	A	A	P	P	P	P	P	P	5
60	MG-U74B	116809_at	unknown	AW120749	A	A	A	P	P	P	P	P	P	5
61	MG-U74B	116839_at	unknown	AK017994	A	M	A	P	P	P	P	P	P	5
62	MG-U74B	116896_at	CGI-49	AF151807	A	A	A	P	P	P	P	P	P	5
63	MG-U74C	128785_r_at	unknown	AI049307	A	A	A	P	P	P	P	P	P	5
64	MG-U74C	128834_r_at	unknown	U66411	A	A	A	P	P	P	P	P	P	5
65	MG-U74C	129843_at	unknown	AA958966	A	A	A	P	P	P	P	P	P	5
66	MG-U74C	130318_at	unknown	AW259417	A	A	A	P	P	P	P	P	P	5

Figure 13C

67	MG-U74C	131255_at	unknown	C88213	A	A	M	P	P	P	P	P	P	5
68	MG-U74C	135403_at	unknown	AK008753	A	A	M	P	M	M	P	P	P	5
69	MG-U74C	136067_at	unknown	AA764539	M	A	M	P	P	P	P	P	P	5
70	MG-U74C	137131_at	DKFZp761A1623	AL137422	A	A	A	P	P	P	P	P	M	5
71	MG-U74C	138548_at	unknown	AJ272344	A	A	A	M	P	P	M	P	P	5
72	MG-U74C	139613_at	unknown	AV258761	M	A	A	P	P	P	P	P	P	5
73	MG-U74C	141159_f_at	unknown	AK009357	M	A	M	M	P	P	M	P	P	5

Figure 14A

	chip	probe set	Gene	Accession number	f1	f2	f3	a1	a2	a3	m1	m2	m.
1	MG-U74A	92581_at	Acetyl coenzyme A dehydrogenase, medium chain	U07159	NC	NC	NC	I	I	I	I	I	I
2	MG-U74A	92587_at	adrenodoxin mRNA	L29123	NC	NC	NC	I	I	I	I	I	I
3	MG-U74A	93119_at	mitochondrial cytochrome c oxidase subunit Vb	X53157	NC	NC	NC	I	I	I	I	I	I
4	MG-U74A	93278_at	sterol-carrier protein 2	M91458	NC	NC	NC	I	I	I	I	I	I
5	MG-U74A	93500_at	amino levulinate synthase	M63245	NC	NC	NC	I	I	I	I	I	I
6	MG-U74A	93572_at	Sterol carrier protein 2	M91458	NC	NC	NC	I	I	I	I	I	I
7	MG-U74A	93581_at	amino levulinate synthase	M63245	NC	NC	NC	I	I	I	I	I	I
8	MG-U74A	93836_at	E1B 19K/Bcl-2-binding protein	AF041054	NC	NC	NC	I	I	I	I	I	I
9	MG-U74A	93972_at	Nadh-Ubiquinone Oxidoreductase 49 kDa Precursor	AF013160	NC	NC	NC	I	I	I	I	I	I
10	MG-U74A	93991_at	Malate dehydrogenase	X07295	NC	NC	NC	I	I	I	I	I	I
11	MG-U74A	94062_at	24-kDa subunit of mitochondrial NADH dehydrogenase	M22756	NC	NC	NC	I	I	I	I	I	I
12	MG-U74A	94078_at	ubiquinol-cytochrome c reductase complex (7.2 kDa)	NM_013387	NC	NC	NC	I	I	I	I	I	I
13	MG-U74A	94216_at	succinate dehydrogenase complex, subunit C,	NM_003001	NC	NC	NC	I	I	I	I	I	I
14	MG-U74A	94375_at	hexokinase II	Y11666	NC	NC	NC	I	I	I	I	I	I
15	MG-U74A	94507_at	long chain fatty acyl CoA synthetase	U15977	NC	NC	NC	I	I	I	I	I	I
16	MG-U74A	94806_at	pyruvate dehydrogenase (lipoamide) beta	BC000439	NC	NC	NC	I	I	I	I	I	I
17	MG-U74A	95053_s_at	succinate dehydrogenase iron-protein subunit	U17248	NC	NC	NC	I	I	I	I	I	I
18	MG-U74A	95064_at	3-oxoacyl-CoA thiolase	X05341	NC	NC	NC	I	I	I	I	I	I
19	MG-U74A	95072_at	Cytochrome C1	X06994	NC	NC	NC	I	I	I	I	I	I
20	MG-U74A	95131_f_at	NADH dehydrogenase 1 beta subcomplex, 2	BC001168	NC	NC	NC	I	I	I	I	I	I
21	MG-U74A	95426_at	mitochondrial enoyl-CoA hydratase	X15958	NC	NC	NC	I	I	I	MI	MI	M
22	MG-U74A	95485_at	3-hydroxyacyl CoA dehydrogenase,	D29639	NC	NC	NC	I	I	I	I	I	I
23	MG-U74A	95646_at	Carnitine palmitoyltransferase 2	U01170	NC	NC	NC	I	I	I	I	I	I
24	MG-U74A	96035_at	branched chain alpha ketoacid decarboxylase E1a subunit	L47335	NC	NC	NC	I	I	I	I	I	I
25	MG-U74A	96112_at	electron-transfer-flavoprotein, alpha	BC003432	NC	NC	NC	I	I	I	I	I	I
26	MG-U74A	96256_at	Anti-oxidant protein 1	M28723	NC	NC	NC	I	I	I	I	I	I

Figure 14B

27	MG-U74A	96267_at	NADH dehydrogenase (ubiquinone) flavoprotein 1	NM_007103	NC	NC	NC	I	I	I	I	I	I
28	MG-U74A	96268_at	succinyl-CoA synthetase	AF144101	NC	NC	NC	I	I	I	I	I	I
29	MG-U74A	96626_at	mitochondrial elongation factor Tu	L38996	NC	NC	NC	MI	I	I	I	I	I
30	MG-U74A	96650_at	AU-binding enoyl-CoA hydratase	AF118386	NC	NC	NC	I	I	I	I	I	I
31	MG-U74A	96746_at	dihydrolipoamide acetyltransferase	D10655	NC	NC	NC	I	I	I	I	I	I
32	MG-U74A	96757_at	homolog of ES1 protein	BC002370	NC	NC	NC	I	I	I	I	I	I
33	MG-U74A	96870_at	aconitase 2, mitochondrial	NM_001098	NC	NC	NC	I	I	I	I	I	I
34	MG-U74A	96899_at	NADH:ubiquinone oxidoreductase NDUFS3 subunit	AF067139	NC	NC	NC	I	I	I	I	I	I
35	MG-U74A	96913_at	mitochondrial long-chain 3-ketoacyl-CoA thiolase	D16479	NC	NC	NC	I	I	I	I	I	I
36	MG-U74A	96947_at	electron transfer flavoprotein	AJ224002	NC	NC	NC	I	I	I	I	I	I
37	MG-U74A	97502_at	Dihydrolipoamide dehydrogenase	U73445	NC	NC	NC	I	I	I	I	I	I
38	MG-U74A	97869_at	electron transfer flavoprotein-ubiquinone oxidoreductase	S69232	NC	NC	NC	I	I	I	I	I	I
39	MG-U74A	97880_at	dihydrolipoamide succinyltransferase	D90401	NC	NC	NC	I	I	I	I	I	I
40	MG-U74A	98132_at	Cytochrome c	U10115	NC	NC	NC	I	I	I	I	I	I
41	MG-U74A	98527_at	dodecenoyl-CoA delta-isomerase	Z14049	NC	NC	NC	I	I	I	I	I	I
42	MG-U74A	98966_at	nuclear-encoded mitochondrial acyltransferase mRNA	L42996	NC	NC	NC	I	I	I	I	I	I
43	MG-U74A	99148_at	mitochondrial fumarase	J04473	NC	NC	NC	I	I	I	I	I	I
44	MG-U74A	99365_at	Dihydroxypolyprenylbenzoate Methyltransferase	AA734487	NC	NC	NC	I	I	I	I	I	I
45	MG-U74A	99613_at	Methylmalonyl coenzyme A mutase	X51941	NC	NC	NC	I	I	I	I	I	I
46	MG-U74A	101525_at	NADH dehydrogenase 1 beta subcomplex, 10	JE0381	NC	NC	NC	I	I	I	I	I	I
47	MG-U74A	101557_f_at	branched chain alpha ketoacid dehydrogenase kinase	AF043070	NC	NC	NC	I	I	I	I	I	I
48	MG-U74A	101867_at	glycerol-3-phosphate acyltransferase	U11680	NC	NC	NC	I	I	I	I	I	I
49	MG-U74A	101989_at	cytochrome-c reductase core I protein	L16842	NC	NC	NC	I	I	I	I	I	I
50	MG-U74A	102000_f_at	ubiquinol-cytochrome c reductase core protein II	BC000484	NC	NC	NC	I	I	I	I	I	I
51	MG-U74A	103646_at	Carnitine acetyltransferase	X85983	NC	NC	NC	I	I	I	I	I	I
52	MG-U74B	106267_s_at	alpha-propionyl-CoA carboxylase	AAA88512	NC	NC	NC	I	I	I	I	I	I
53	MG-U74B	106268_at	alpha-propionyl-CoA carboxylase	AAA88512	NC	NC	NC	I	I	I	I	I	I
54	MG-U74B	111690_at	pyruvate dehydrogenase (lipoamide) beta	BC000439	NC	NC	NC	I	I	I	I	I	I

Figure 14C

55	*MG-U74B*	*112974_at*	*energy-dependent regulator of proteolysis*	*AF134983*	NC	NC	NC	I	I	I	I	I	I
56	*MG-U74B*	*115077_f_at*	*dihydrolipoamide branched chain transacylase E2*	*NM_010022*	NC	NC	NC	I	I	I	I	I	I
57	*MG-U74B*	*115211_at*	*dihydrolipoamide acetyl transferase*	*U79296*	NC	NC	NC	I	I	I	I	I	I
58	*MG-U74B*	*115357_at*	*sodium/hydrogen exchanger isoform 6*	*XM_010374*	NC	NC	NC	I	I	I	I	I	I
59	MG-U74A	AFFX-GapdhMur/	glyceraldehyde-3-phosphate dehydrogenase	M32599	NC	NC	NC	I	I	I	I	I	I
60	MG-U74A	92592_at	Glycerolphosphate dehydrogenase 1, cytoplasmic	M25558	NC	NC	NC	I	I	I	I	I	I
61	MG-U74A	93029_at	Isocitrate dehydrogenase 3 (NAD+), gamma	U68564	NC	NC	NC	I	I	I	I	I	I
62	MG-U74A	94056_at	Stearoyl-coenzyme A desaturase 1	M21285	NC	NC	NC	I	I	I	I	I	I
63	MG-U74A	94057_g_at	Stearoyl-coenzyme A desaturase 1	M21285	NC	NC	NC	I	I	I	I	I	I
64	MG-U74A	96095_i_at	Estradiol 17 Beta-Dehydrogenase 4	AI648018	NC	NC	NC	MI	MI	I	I	I	I
65	MG-U74A	96321_at	similar to NADH dehydrogenase 1 alpha subcomplex, 9	BC003351	NC	NC	NC	I	I	I	I	I	I
66	MG-U74A	96348_at	malate NADP oxidoreductase	J02652	NC	NC	NC	I	I	I	I	I	I
67	MG-U74A	98575_at	Fatty acid synthase	X13135	NC	NC	NC	I	I	I	I	I	I
68	MG-U74A	98909_at	putative lipoic acid synthetase	AJ224162	NC	NC	NC	I	I	I	I	I	I
69	MG-U74A	99481_at	ATPase, Na+K+ transporting, alpha 2	NM_012505	NC	NC	NC	I	I	I	I	I	I
70	MG-U74A	100042_at	hydroxyacyl glutathione hydrolase	NM_005326	NC	NC	NC	I	I	I	I	I	I
71	MG-U74A	101044_at	Delta-aminolevulinate dehydratase	X13752	NC	NC	NC	I	I	I	I	I	I
72	MG-U74A	101082_at	Malic enzyme, supernatant	J02652	NC	NC	NC	I	I	I	I	I	I
73	MG-U74A	101991_at	flavin-containing monooxygenase	D16215	NC	NC	NC	I	I	I	I	I	I
74	MG-U74A	102022_at	unknown Glyoxalase	AK003511	NC	NC	NC	I	I	I	I	I	I
75	MG-U74A	102208_at	alpha-2,3-sialyltransferase VI	AF119390	NC	NC	NC	I	I	I	I	I	I
76	MG-U74A	102314_at	Glut4	M23383	NC	NC	NC	I	I	I	I	I	I
77	MG-U74A	103087_at	phenol/aryl form sulfotransferase	L02331	NC	NC	NC	I	I	I	MI	MI	M
78	MG-U74A	103297_at	6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase	X98848	NC	NC	NC	I	I	I	I	I	I
79	MG-U74A	104313_at	phosphoglucomutase 1	M83088	NC	NC	NC	I	I	I	I	I	I
80	MG-U74B	107117_at	peptide methionine sulfoxide reductase	AJ242973	NC	NC	NC	I	I	I	I	I	I
81	MG-U74B	108537_at	similar to putative type III alcohol dehydrogenase	AAB71082	NC	NC	NC	I	I	I	I	I	I
82	MG-U74B	113545_at	ATPase, Na+K+ transporting, alpha 2	NM_012505	NC	NC	NC	I	I	I	I	I	I

Figure 14D

83	MG-U74B	113915_at	glutamic-pyruvate transaminase	XM_011679	NC	NC	NC	I	I	I	I	I	I
84	MG-U74C	132388_f_at	alanine aminotransferase	D10354	NC	NC	NC	I	I	I	I	I	I
85	MG-U74A	92202_g_at	Q05516 zinc finger protein plzf	AI553024	NC	NC	NC	I	I	I	I	I	I
86	MG-U74A	92537_g_at	beta-3-adrenergic receptor	X72862	NC	NC	NC	I	I	I	I	I	M
87	MG-U74A	93082_at	AWP1	AJ251508	NC	NC	NC	I	I	I	I	I	I
88	MG-U74A	94365_at	Histidine Triad Nucleotide-Binding Protein	AA710175	NC	NC	NC	I	I	I	I	I	I
89	MG-U74A	95074_at	Peroxisomal farnesylated protein	NM_023041	NC	NC	NC	MI	I	MI	I	I	I
90	MG-U74A	97248_at	diazepam-binding inhibitor	X61431	NC	NC	NC	I	I	I	I	I	I
91	MG-U74A	98428_at	spastin protein	AJ246002	NC	NC	NC	I	MI	I	MI	I	I
92	MG-U74A	98892_at	Lpin1	AF180471	NC	NC	NC	I	I	I	I	I	I
93	MG-U74A	99532_at	Tob1	D78382	NC	NC	NC	I	I	I	I	I	I
94	MG-U74A	102366_at	Resisitn	AA718169	NC	NC	NC	I	I	I	I	I	I
95	MG-U74A	103015_at	B-cell leukemia/lymphoma 6	U41465	NC	NC	NC	I	I	I	I	I	I
96	MG-U74A	103833_at	homeodomain-interacting protein kinase 2	AF077659	NC	NC	NC	I	I	I	I	I	I
97	MG-U74A	104342_i_at	AKR group XII-1 phospholipase A2	AY007381	NC	NC	NC	I	I	I	I	I	I
98	MG-U74A	104343_f_at	AKR group XII-1 phospholipase A2	AY007381	NC	NC	NC	I	I	I	I	I	I
99	MG-U74A	104745_at	ADP-ribosylation-like factor 6 interacting protein 2	NM_019717	NC	NC	NC	I	I	I	I	I	I
100	MG-U74B	106573_at	SNF1-related kinase	X89383	NC	NC	NC	I	I	I	I	I	M
101	MG-U74B	106577_at	BTEB-1 transcription factor.	Y14296	NC	NC	NC	I	I	I	I	I	I
102	MG-U74B	109110_at	protein phosphatase 2C beta-1	AJ271836	NC	NC	NC	I	I	I	I	I	I
103	MG-U74B	110650_at	nuclear receptor interacting protein 1	XM_009699	NC	NC	NC	I	I	I	I	I	I
104	MG-U74B	111708_at	Lpin1	AF180471	NC	NC	NC	I	I	I	I	I	I
105	MG-U74B	111784_at	calcium-independent phospholipase A2	AB041261	NC	NC	NC	I	I	I	I	I	I
106	MG-U74B	112336_at	Unknown seven transmembrane receptor	AB019120	NC	NC	NC	I	I	I	I	I	I
107	MG-U74B	113124_at	Kruppel-like transcription factor	AB020759	NC	NC	NC	I	I	I	I	I	I
108	MG-U74B	113762_at	pre-B-cell colony-enhancing factor	AF234625	NC	NC	NC	I	I	I	I	I	I
109	MG-U74B	113976_at	mitogen-activated protein kinase phosphatase x	XM_004537	NC	NC	NC	I	I	I	I	I	I
110	MG-U74B	114062_at	kinase interacting protein 2	NM_019686	NC	NC	NC	I	I	I	I	I	I

Figure 14E

111	MG-U74C	130718_at	tissue inhibitor of metalloproteinases TIMP4	U76456	NC	NC	NC	I	I	I	I	I	I
112	MG-U74A	94269_at	rab6/rab5-associated protein	L40934	NC	NC	NC	I	I	I	I	I	I
113	MG-U74A	99111_at	SKD3	U09874	NC	NC	NC	I	I	I	I	I	I
114	MG-U74B	107534_at	tensin	NM_022648	NC	NC	NC	MI	I	I	I	I	I
115	MG-U74B	108055_at	tensin	NM_022648	NC	NC	NC	I	I	I	I	I	I
116	MG-U74B	111431 at	alpha-actinin 1	M95178	NC	NC	NC	I	I	I	I	I	I
117	MG-U74A	92805_s_at	ADP-ribosylation-like 4	Y12577	NC	NC	NC	I	I	I	I	I	I
118	MG-U74A	95453_f_at	calcium-binding protein S100A1	AF087687	NC	NC	NC	I	I	I	I	I	I
119	MG-U74A	95736_at	Unknown protein Ribosomal protein L4/L1 family	AB041599	NC	NC	NC	I	I	I	I	I	I
120	MG-U74A	95787_s_at	sterol carrier protein 2-pseudogene	X87685	NC	NC	NC	I	I	I	I	I	I
121	MG-U74A	96296_at	Ribosomal protein L15	AK010604	NC	NC	NC	I	MI	I	I	I	I
122	MG-U74A	96684_at	G-Rich Sequence Factor-1	NM_002092	NC	NC	NC	I	I	I	I	I	I
123	MG-U74A	97204_s_at	MCJ	AF126743	NC	NC	NC	I	I	I	I	I	I
124	MG-U74A	99104_at	adipoQ	U49915	NC	NC	NC	I	I	I	I	I	I
125	MG-U74A	99115_at	calcium binding protein A11	NM_009112	NC	NC	NC	I	I	I	I	I	I
126	MG-U74A	99671_at	adipsin	L42996	NC	NC	NC	I	I	I	I	I	I
127	MG-U74A	100567_at	adipocyte P2	M20497	NC	NC	NC	I	I	I	I	I	I
128	MG-U74A	100596_at	Selenium-binding liver protein	M32032	NC	NC	NC	I	I	I	I	I	I
129	MG-U74B	109884_f_at	CD36	L23108	NC	NC	NC	I	I	I	I	I	I
130	MG-U74B	115504_g_at	dysferlin	AJ242954	NC	NC	NC	I	I	I	I	I	I
131	MG-U74A	92542_at	RIKEN cDNA 2700043I21 gene	NM_023665	NC	NC	NC	I	I	I	I	I	I
132	MG-U74A	92941_at	Unknown	AK004252	NC	NC	NC	MI	MI	I	I	I	I
133	MG-U74A	94327_at	FLJ10548	AK001410	NC	NC	NC	I	I	I	I	I	I
134	MG-U74A	94531_at	HSPC326	AF161444	NC	NC	NC	I	I	I	I	I	I
135	MG-U74A	95058_f_at	brain 0-44	M13095	NC	NC	NC	I	I	I	I	I	I
136	MG-U74A	95119_at	HSPC326	AF161444	NC	NC	NC	I	I	I	I	I	I
137	MG-U74A	95468_at	unknown	none	NC	NC	NC	I	I	I	I	I	I
138	MG-U74A	95690_at	CGI-51 protein	XM_010037	NC	NC	NC	I	I	I	I	I	I

Figure 14F

139	MG-U74A	96096_f_at	MGC:7467	AI648018	NC	NC	NC	I	I	I	I	I	I
140	MG-U74A	96122_at	3 TCP-1 epsilon	AW049373	NC	NC	NC	I	I	I	I	I	I
141	MG-U74A	96158_at	KIAA1155	AB032981	NC	NC	NC	I	I	I	I	I	I
142	MG-U74A	96212_at	unknown	AAF47115	NC	NC	NC	I	I	I	I	I	I
143	MG-U74A	96237_at	unknown	AI118905	NC	NC	NC	I	I	I	I	I	I
144	MG-U74A	96615_at	unknown	AK002925	NC	NC	NC	I	I	I	I	MI	I
145	MG-U74A	96680_at	mDj7	AB028857	NC	NC	NC	I	I	I	I	I	I
146	MG-U74A	97420_at	Riken clone:1300008B03	AK004940	NC	NC	NC	I	I	MI	I	I	I
147	MG-U74A	97511_at	unknown	AI846600	NC	NC	NC	I	I	I	I	I	I
148	MG-U74A	97908_at	hypothetical protein	AJ237586	NC	NC	NC	I	I	I	I	I	I
149	MG-U74A	98039_at	sr104 protein	AJ250690	NC	NC	NC	I	I	I	I	I	I
150	MG-U74A	99471_at	1-17 unknown mRNA	AF244543	NC	NC	NC	I	I	I	I	I	I
151	MG-U74A	99594_at	unknown	AK012126	NC	NC	NC	I	I	I	I	I	I
152	MG-U74A	99636_at	unknown	AK003702	NC	NC	NC	I	I	I	I	I	I
153	MG-U74A	99849_at	Murine leukemia virus VLeco long terminal repeat	M15425	NC	NC	NC	I	I	I	I	I	I
154	MG-U74A	101094_at	hypoxia induced gene 1	AF141312	NC	NC	NC	I	I	I	I	I	I
155	MG-U74A	101912_at	Unknown	AI01967	NC	NC	NC	I	I	I	I	I	I
156	MG-U74A	102991_s_at	major histocompatibility locus class II region	AF100956	NC	NC	NC	I	I	I	I	I	I
157	MG-U74A	104098_at	peroxisome membrane protein	L28835	NC	NC	NC	I	I	I	I	I	I
158	MG-U74A	104396_at	unknown	AI019327	NC	NC	NC	I	I	I	MI	I	I
159	MG-U74A	104605_at	FLJ21432	AK025085	NC	NC	NC	I	I	I	I	I	I
160	MG-U74B	105993_at	PTD015 protein	BC00272	NC	NC	NC	I	I	I	I	I	I
161	MG-U74B	107059_at	similar to CG15168 gene product	AL157477	NC	NC	NC	I	I	I	I	I	I
162	MG-U74B	107283_at	MNCb-3350	NM_020589	NC	NC	NC	I	I	I	I	I	I
163	MG-U74B	108108_at	hypothalamus protein HT008	XM_008246	NC	NC	NC	MI	I	I	I	I	I
164	MG-U74B	108493_at	ORF1	AJ278735	NC	NC	NC	I	I	I	I	I	I
165	MG-U74B	108749_at	Unknown	AA611885	NC	NC	NC	I	I	I	I	I	I
166	MG-U74B	109008_at	Unknown	AK015530	NC	NC	NC	I	I	I	I	I	I

Figure 14G

167	MG-U74B	109016_at	unknown	AK008857	NC	NC	NC	I	I	I	I	I	I
168	MG-U74B	110471_i_at	KIAA0172	AK025636	NC	NC	NC	I	I	I	I	I	I
169	MG-U74B	110826_at	HSPC202	AF151036	NC	NC	NC	I	I	I	I	I	I
170	MG-U74B	112029_at	unknown	AA821545	NC	NC	NC	I	I	I	I	I	I
171	MG-U74B	112361_at	CGI-10	AF132944	NC	NC	NC	I	I	I	I	I	I
172	MG-U74B	112864_at	FLJ12592	AK022654	NC	NC	NC	I	I	I	I	I	I
173	MG-U74B	113212_at	unknown protein	AK007496	NC	NC	NC	I	I	I	I	I	I
174	MG-U74B	113599_at	Similar to BACR7C10.a [Drosophila melanogaster].	CAB58018	NC	NC	NC	I	I	I	I	I	I
175	MG-U74B	113968_at	unknown	AI849005	NC	NC	NC	I	I	I	I	I	I
176	MG-U74B	114296_at	FLJ10378	AK001240	NC	NC	NC	I	MI	I	I	I	I
177	MG-U74B	116644_at	LRP16	BC000270	NC	NC	NC	I	I	I	I	I	I
178	MG-U74B	117208_at	unknown	AK003388	NC	NC	NC	I	I	I	I	I	I
179	MG-U74B	117301_at	clone 24418	AF052185	NC	NC	NC	I	I	I	I	I	I
180	MG-U74C	129952_at	unknown	AI595378	NC	NC	NC	MI	I	I	I	I	I
181	MG-U74C	131255_at	unknown	C88213	NC	NC	NC	I	I	I	I	I	I
182	MG-U74C	132466_f_at	MGC:7467	BC002211	NC	NC	NC	MI	I	I	I	I	I
183	MG-U74C	135381_i_at	unknown	AW20847	NC	NC	NC	MI	I	I	I	I	I
184	MG-U74C	135382_r_at	unknown	AW20847	NC	NC	NC	MI	I	I	I	I	I
185	MG-U74C	136067_at	unknown	AA764539	NC	NC	NC	MI	I	I	I	I	I
186	MG-U74C	137124_at	unknown	AA124582	NC	NC	NC	I	I	I	I	I	I
187	MG-U74C	141159_f_at	unknown	AK009357	NC	NC	NC	I	I	I	I	I	I

Figure 15A

Highly Expressed Genes common between the Muscle-Adipocyte Union library and the Mu-74 GeneChips Arrays.

	Name	Accession number	Probe sets	Adipocyte increase	Muscle increase
U001	CD36	L23108	93332_at	1.5 -0.5	2.8 -0.8
U003	glyceraldehyde-3-phosphate dehyrdogenase	M32599	U74B AFFX-GapdhMur/M32599_5_at	1.9 -1.0	5.4 -0.3
U003	glyceraldehyde-3-phosphate dehyrdogenase	M32599	U74A AFFX-GapdhMur/M32599_5_at	2.3 -0.9	4.7 -0.3
U004	hexokinase II	Y11666	94375_at	18.5 -18.0	28.1 -3.0
U009	diazepam-binding inhibitor	X61431	97248_at	3.9 0.0	2.3 -0.1
U010	Glycerolphosphate dehydrogenase 1, cytoplasmic	M25558	92592_at	24.1 -2.0	21.3 -1.8
U010	Glycerolphosphate dehydrogenase1, cytoplasmic	M25558	92592_at	24.1 -2.0	21.3 -1.8
U012	Carnitine palmitoyltransferase 2	U01170	95646_at	4.3 -0.1	3.7 -0.1
U014	Glut 4	M23383	102314_at	12.9 -0.4	35.8 -5.4
U015	adipose fatty acid binding protein	M20497	100567_at	67.8 -17.3	21.5 -5.7
U016	Stearoyl-coenzyme A desaturase 1	M21285	94056_at	30.2 -6.7	13.0 -1.9
U016	Stearoyl-coenzyme A desaturase 1	M21285	94057_g_at	114.5 -17.4	22.3 -3.2
U017	AdipoQ	U49915	99104_at	257.2 -70.9	35.7 -9.7
U018	Adipsin	X04673	99671_at	405.4 -103.4	55.5 -10.8
U019	long chain fatty acyl CoA synthetase	U15977	94507_at	35.9 -5.4	13.2 -0.6
U020	Resistin	AA718169	102366_at	109 -43.8	9.7 -2.0

Figure 15B

U029	CAP	AI536451	113344_at	4.8 -1.1	1.2 -0.4
U031	Lpin1	AI846934	98892_at	5.2 -1.1	4.3 -0.3
U031	Lpin1	AA709944	111708_at	10.7 -2.0	17.6 -0.6
U046	pre-B-cell colony enhancing factor	AI510151	113762_at	2.1 -0.8	2.4 -0.1
U047	calcium-binding protein S100A1	AF087687	95453_f_at	30.5 -13.5	10.2 -1.7
U066	ATPase, NA+K+ transporting, alpha 2	AI847141	113545_at	2.7 -1.0	25.1 -1.6
U074	mitochondrial enoyl-CoA hydratase	AW048512	95426_at	11.3 -3.4	4.3 -0.9
U078	24-kDa subunit of mitochondrial NADH dehydrogenase	AI847609	94062_at	2.8 -0.7	6.6 -0.2
U096	NADH	X06994	95072_at	2.4 -0.5	4.0 -0.2
U098	flavin-containing monooxygenase	D16215	101991_at	4.4 -1.7	4.7 -1.1
U106	NADH dehydrogenase	Y07708	98118_at	0.9 -0.4	0.6 -0.0
U111	phenol/aryl form sulfotransferase	L02331	103087_at	12.7 -7.4	2.5 -0.1
U195	CGI-51 protein	AW047363	95690_at	2.2 -0.5	3.9 -0.3
U214	Unknown	AI846600	97511_at	5.5 -1.6	10.6 -1.2
U226	unknown weakly similar to putative type III alcohol dehydrogenase	AI846354	108537_at	18.4 -4.5	5.9 -0.1
U226	unknown weakly similar to putative type III alcohol dehydrogenase	AA710501	128834_r_at	3.6 -0.5	2.0 -1.0